

1. Biology

1. [Energy and Metabolism](#)
2. [Potential, Kinetic, Free, and Activation Energy](#)
3. [The Laws of Thermodynamics](#)
4. [ATP: Adenosine Triphosphate](#)
5. [Enzymes](#)
6. [Energy in Living Systems](#)
7. [Glycolysis](#)
8. [Oxidative Phosphorylation](#)
9. [Metabolism without Oxygen](#)
10. [Connections of Carbohydrate, Protein, and Lipid Metabolic Pathways](#)
11. [Regulation of Cellular Respiration](#)
12. [Overview of Photosynthesis](#)
13. [The Light-Dependent Reactions of Photosynthesis](#)
14. [Using Light Energy to Make Organic Molecules](#)
15. [Nutrition and Energy Production](#)
16. [Muscle Contraction and Locomotion](#)
17. [Energy Flow through Ecosystems](#)
18. [Biogeochemical Cycles](#)

2. Chemistry

1. [Energy Basics](#)
2. [Calorimetry](#)
3. [Enthalpy](#)
4. [Electromagnetic Energy](#)
5. [The Bohr Model](#)
6. [Development of Quantum Theory](#)
7. [Electronic Structure of Atoms \(Electron Configurations\)](#)
8. [Periodic Variations in Element Properties](#)
9. [Chemical Reaction Rates](#)
10. [Spontaneity](#)

11. [Entropy](#)
12. [The Second and Third Laws of Thermodynamics](#)
13. [Free Energy](#)
14. [Electrode and Cell Potentials](#)
15. [Potential, Free Energy, and Equilibrium](#)
16. [Batteries and Fuel Cells](#)
17. [Corrosion](#)
18. [Electrolysis](#)
3. Earth and Space
 1. [The Behavior of Light](#)
 2. [The Electromagnetic Spectrum](#)
 3. [Spectroscopy in Astronomy](#)
 4. [The Structure of the Atom](#)
 5. [Formation of Spectral Lines](#)
 6. [The Doppler Effect](#)
 7. [Earth's Crust](#)
 8. [Earth's Atmosphere](#)
 9. [Life, Chemical Evolution, and Climate Change](#)
 10. [The Geology of Venus](#)
 11. [Atmospheres of the Giant Planets](#)
 12. [The Structure and Composition of the Sun](#)
 13. [The Solar Cycle](#)
 14. [Solar Activity above the Photosphere](#)
 15. [Space Weather](#)
 16. [Sources of Sunshine: Thermal and Gravitational Energy](#)
 17. [Mass, Energy, and the Theory of Relativity](#)
 18. [The Solar Interior: Theory](#)
 19. [The Solar Interior: Observations](#)
4. Physics
 1. [Work: The Scientific Definition](#)
 2. [Kinetic Energy and the Work-Energy Theorem](#)
 3. [Gravitational Potential Energy](#)

4. [Conservative Forces and Potential Energy](#)
5. [Nonconservative Forces](#)
6. [Conservation of Energy](#)
7. [Power](#)
8. [Work, Energy, and Power in Humans](#)
9. [World Energy Use](#)
10. [Temperature](#)
11. [Thermal Expansion of Solids and Liquids](#)
12. [The Ideal Gas Law](#)
13. [Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature](#)
14. [Phase Changes](#)
15. [Humidity, Evaporation, and Boiling](#)
16. [Heat](#)
17. [Temperature Change and Heat Capacity](#)
18. [Phase Change and Latent Heat](#)
19. [Heat Transfer Methods](#)
20. [Conduction](#)
21. [Convection](#)
22. [Radiation](#)
23. [The First Law of Thermodynamics](#)
24. [The First Law of Thermodynamics and Some Simple Processes](#)
25. [Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency](#)
26. [Carnot's Perfect Heat Engine: The Second Law of Thermodynamics Restated](#)
27. [Applications of Thermodynamics: Heat Pumps and Refrigerators](#)
28. [Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy](#)

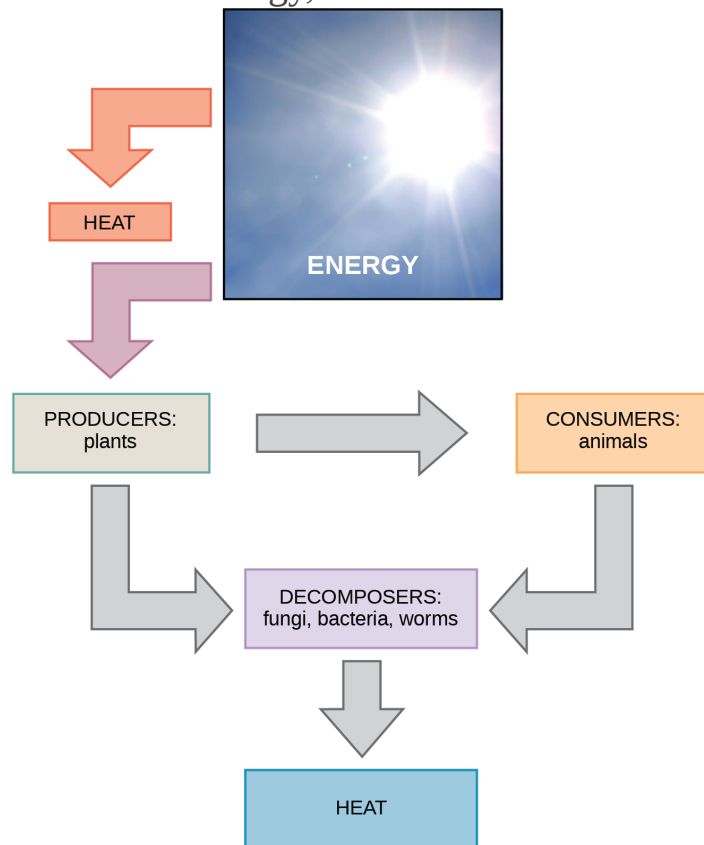
29. [Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation](#)
30. [Energy and the Simple Harmonic Oscillator](#)
31. [Waves](#)
32. [Energy in Waves: Intensity](#)
33. [Maxwell's Equations: Electromagnetic Waves Predicted and Observed](#)
34. [Production of Electromagnetic Waves](#)
35. [The Electromagnetic Spectrum](#)
36. [Energy in Electromagnetic Waves](#)
37. [Quantization of Energy](#)
38. [The Photoelectric Effect](#)
39. [Photon Energies and the Electromagnetic Spectrum](#)
40. [Photon Momentum](#)
41. [The Particle-Wave Duality](#)
42. [The Wave Nature of Matter](#)
43. [Nuclear Radioactivity](#)
44. [Binding Energy](#)
45. [Tunneling](#)

Energy and Metabolism

By the end of this section, you will be able to do the following:

- Explain metabolic pathways and describe the two major types
- Discuss how chemical reactions play a role in energy transfer

Scientists use the term **bioenergetics** to discuss the concept of energy flow ([link](#)) through living systems, such as cells. Cellular processes such as building and breaking down complex molecules occur through stepwise chemical reactions. Some of these chemical reactions are spontaneous and release energy; whereas, others require energy to proceed. Just as living things must continually consume food to replenish what they have used, cells must continually obtain more energy to replenish that which the many energy-requiring chemical reactions that constantly take place use. All of the chemical reactions that transpire inside cells, including those that use and release energy, are the cell's **metabolism**.



Most life forms on earth obtain their energy from the sun. Plants use

photosynthesis to capture sunlight, and herbivores eat those plants to obtain energy. Carnivores eat the herbivores, and decomposers digest plant and animal matter.

Carbohydrate Metabolism

Sugar (chemical reactions) metabolism (a simple carbohydrate) is a classic example of the many cellular processes that use and produce energy. Living things consume sugar as a major energy source, because sugar molecules have considerable energy stored within their bonds. The following equation describes the breakdown of glucose, a simple sugar:

Equation:



Consumed carbohydrates have their origins in photosynthesizing organisms like plants ([link](#)). During photosynthesis, plants use the energy of sunlight to convert carbon dioxide gas (CO_2) into sugar molecules, like glucose ($\text{C}_6\text{H}_{12}\text{O}_6$). Because this process involves synthesizing a larger, energy-storing molecule, it requires an energy input to proceed. The following equation (notice that it is the reverse of the previous equation) describes the synthesis of glucose:

Equation:



During photosynthesis chemical reactions, energy is in the form of a very high-energy molecule scientists call ATP, or adenosine triphosphate. This is the primary energy currency of all cells. Just as the dollar is the currency we use to buy goods, cells use ATP molecules as energy currency to perform immediate work. The sugar (glucose) is stored as starch or glycogen.

Energy-storing polymers like these break down into glucose to supply ATP molecules.

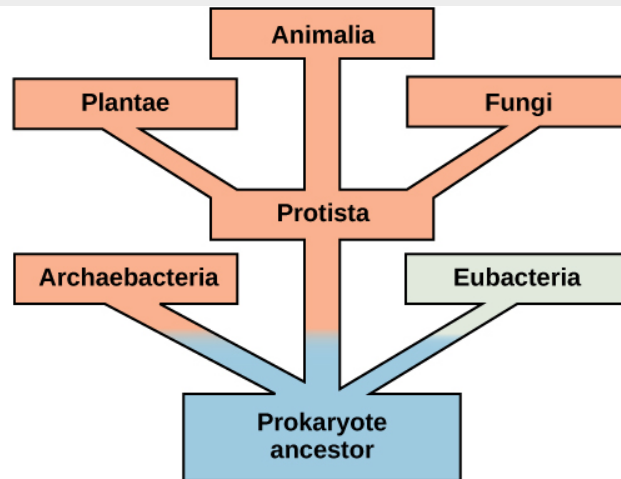
Solar energy is required to synthesize a glucose molecule during the photosynthesis reactions. In photosynthesis, light energy from the sun initially transforms into chemical energy that temporally stores itself in the energy carrier molecules ATP and NADPH (nicotinamide adenine dinucleotide phosphate). Photosynthesis later uses the stored energy in ATP and NADPH to build one glucose molecule from six molecules of CO_2 . This process is analogous to eating breakfast in the morning to acquire energy for your body that you can use later in the day. Under ideal conditions, energy from 18 molecules of ATP is required to synthesize one glucose molecule during photosynthesis reactions. Glucose molecules can also combine with and convert into other sugar types. When an organism consumes sugars, glucose molecules eventually make their way into each organism's living cell. Inside the cell, each sugar molecule breaks down through a complex series of chemical reactions. The goal of these reactions is to harvest the energy stored inside the sugar molecules. The harvested energy makes high-energy ATP molecules, which perform work, powering many chemical reactions in the cell. The amount of energy needed to make one glucose molecule from six carbon dioxide molecules is 18 ATP molecules and 12 NADPH molecules (each one of which is energetically equivalent to three ATP molecules), or a total of 54 molecule equivalents required for synthesizing one glucose molecule. This process is a fundamental and efficient way for cells to generate the molecular energy that they require.



Plants, like this oak tree and acorn, use energy from sunlight to make sugar and other organic molecules. Both plants and animals (like this squirrel) use cellular respiration to derive energy from the organic molecules that plants originally produced. (credit “acorn”: modification of work by Noel Reynolds; credit “squirrel”: modification of work by Dawn Huczek)

Metabolic Pathways

The processes of making and breaking down sugar molecules illustrate two types of metabolic pathways. A metabolic pathway is a series of interconnected biochemical reactions that convert a substrate molecule or molecules, step-by-step, through a series of metabolic intermediates, eventually yielding a final product or products. In the case of sugar metabolism, the first metabolic pathway synthesized sugar from smaller molecules, and the other pathway broke sugar down into smaller molecules. Scientists call these two opposite processes—the first requiring energy and the second producing energy—anabolic (building) and catabolic (breaking down) pathways, respectively. Consequently, building (anabolism) and degradation (catabolism) comprise metabolism.

Note:**Evolution Connection**

This tree shows the evolution of the various branches of life. The vertical dimension is time. Early life forms, in blue, used anaerobic metabolism to obtain energy from their surroundings.

Evolution of Metabolic Pathways

There is more to the complexity of metabolism than understanding the metabolic pathways alone. Metabolic complexity varies from organism to organism. Photosynthesis is the primary pathway in which photosynthetic organisms like plants (planktonic algae perform the majority of global synthesis) harvest the sun's energy and convert it into carbohydrates. The by-product of photosynthesis is oxygen, which some cells require to carry out cellular respiration. During cellular respiration, oxygen aids in the catabolic breakdown of carbon compounds, like carbohydrates. Among the products are CO_2 and ATP. In addition, some eukaryotes perform catabolic processes without oxygen (fermentation); that is, they perform or use anaerobic metabolism.

Organisms probably evolved anaerobic metabolism to survive (living organisms came into existence about 3.8 billion years ago, when the atmosphere lacked oxygen). Despite the differences between organisms and the complexity of metabolism, researchers have found that all branches of life share some of the same metabolic pathways, suggesting that all

organisms evolved from the same ancient common ancestor ([\[link\]](#)). Evidence indicates that over time, the pathways diverged, adding specialized enzymes to allow organisms to better adapt to their environment, thus increasing their chance to survive. However, the underlying principle remains that all organisms must harvest energy from their environment and convert it to ATP to carry out cellular functions.

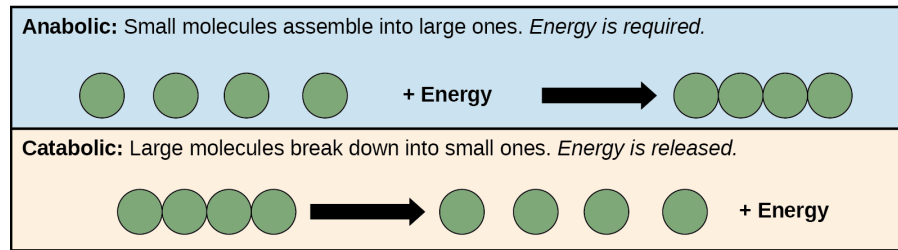
Anabolic and Catabolic Pathways

Anabolic pathways require an input of energy to synthesize complex molecules from simpler ones. Synthesizing sugar from CO_2 is one example. Other examples are synthesizing large proteins from amino acid building blocks, and synthesizing new DNA strands from nucleic acid building blocks. These biosynthetic processes are critical to the cell's life, take place constantly, and demand energy that ATP and other high-energy molecules like NADH (nicotinamide adenine dinucleotide) and NADPH provide ([\[link\]](#)).

ATP is an important molecule for cells to have in sufficient supply at all times. The breakdown of sugars illustrates how a single glucose molecule can store enough energy to make a great deal of ATP, 36 to 38 molecules. This is a **catabolic** pathway. Catabolic pathways involve degrading (or breaking down) complex molecules into simpler ones. Molecular energy stored in complex molecule bonds release in catabolic pathways and harvest in such a way that it can produce ATP. Other energy-storing molecules, such as fats, also break down through similar catabolic reactions to release energy and make ATP ([\[link\]](#)).

It is important to know that metabolic pathway chemical reactions do not take place spontaneously. A protein called an enzyme facilitates or catalyzes each reaction step. Enzymes are important for catalyzing all types of biological reactions—those that require energy as well as those that release energy.

Metabolic pathways



Anabolic pathways are those that require energy to synthesize larger molecules. Catabolic pathways are those that generate energy by breaking down larger molecules. Both types of pathways are required for maintaining the cell's energy balance.

Section Summary

Cells perform the functions of life through various chemical reactions. A cell's metabolism refers to the chemical reactions that take place within it. There are metabolic reactions that involve breaking down complex chemicals into simpler ones, such as breaking down large macromolecules. Scientists refer to this process as catabolism, and we associate such reactions an energy release. On the other end of the spectrum, anabolism refers to metabolic processes that build complex molecules out of simpler ones, such as macromolecule synthesis. Anabolic processes require energy. Glucose synthesis and glucose breakdown are examples of anabolic and catabolic pathways, respectively.

Multiple Choice

Exercise:

Problem:

Energy is stored long-term in the bonds of _____ and used short-term to perform work from a(n) _____ molecule.

- a. ATP : glucose
 - b. an anabolic molecule : catabolic molecule
 - c. glucose : ATP
 - d. a catabolic molecule : anabolic molecule
-

Solution:

C

Exercise:

Problem:

DNA replication involves unwinding two strands of parent DNA, copying each strand to synthesize complementary strands, and releasing the parent and daughter DNA. Which of the following accurately describes this process?

- a. This is an anabolic process.
 - b. This is a catabolic process.
 - c. This is both anabolic and catabolic.
 - d. This is a metabolic process but is neither anabolic nor catabolic.
-

Solution:

A

Critical Thinking Questions

Exercise:

Problem:

Does physical exercise involve anabolic and/or catabolic processes?
Give evidence for your answer.

Solution:

Physical exercise involves both anabolic and catabolic processes. Body cells break down sugars to provide ATP to do the work necessary for exercise, such as muscle contractions. This is catabolism. Muscle cells also must repair muscle tissue damaged by exercise by building new muscle. This is anabolism.

Exercise:

Problem:

Name two different cellular functions that require energy that parallel human energy-requiring functions.

Solution:

Energy is required for cellular motion, through beating of cilia or flagella, as well as human motion, produced by muscle contraction. Cells also need energy to perform digestion, as humans require energy to digest food.

Glossary

anabolic

(also, anabolism) pathways that require an energy input to synthesize complex molecules from simpler ones

bioenergetics

study of energy flowing through living systems

catabolic

(also, catabolism) pathways in which complex molecules break down into simpler ones

metabolism

all the chemical reactions that take place inside cells, including anabolism and catabolism

Potential, Kinetic, Free, and Activation Energy

By the end of this section, you will be able to do the following:

- Define “energy”
- Explain the difference between kinetic and potential energy
- Discuss the concepts of free energy and activation energy
- Describe endergonic and exergonic reactions

We define energy as the ability to do work. As you’ve learned, energy exists in different forms. For example, electrical energy, light energy, and heat energy are all different energy types. While these are all familiar energy types that one can see or feel, there is another energy type that is much less tangible. Scientists associate this energy with something as simple as an object above the ground. In order to appreciate the way energy flows into and out of biological systems, it is important to understand more about the different energy types that exist in the physical world.

Energy Types

When an object is in motion, there is energy. For example, an airplane in flight produces considerable energy. This is because moving objects are capable of enacting a change, or doing work. Think of a wrecking ball. Even a slow-moving wrecking ball can do considerable damage to other objects. However, a wrecking ball that is not in motion is incapable of performing work. Energy with objects in motion is **kinetic energy**. A speeding bullet, a walking person, rapid molecule movement in the air (which produces heat), and electromagnetic radiation like light all have kinetic energy.

What if we lift that same motionless wrecking ball two stories above a car with a crane? If the suspended wrecking ball is unmoving, can we associate energy with it? The answer is yes. The suspended wrecking ball has associated energy that is fundamentally different from the kinetic energy of objects in motion. This energy form results from the *potential* for the wrecking ball to do work. If we release the ball it would do work. Because this energy type refers to the potential to do work, we call it **potential energy**. Objects transfer their energy between kinetic and potential in the

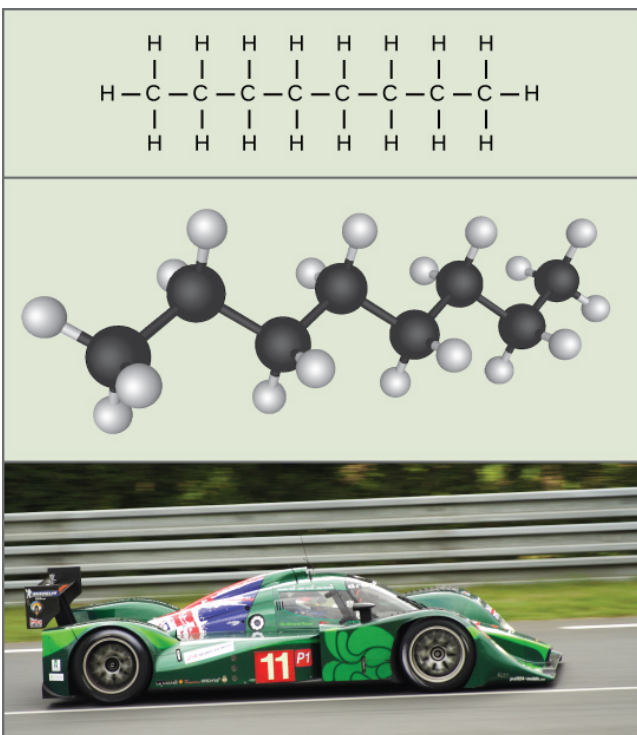
following way: As the wrecking ball hangs motionless, it has 0 kinetic and 100 percent potential energy. Once it releases, its kinetic energy begins to increase because it builds speed due to gravity. Simultaneously, as it nears the ground, it loses potential energy. Somewhere mid-fall it has 50 percent kinetic and 50 percent potential energy. Just before it hits the ground, the ball has nearly lost its potential energy and has near-maximal kinetic energy. Other examples of potential energy include water's energy held behind a dam ([link](#)), or a person about to skydive from an airplane.



Water behind a dam has potential energy. Moving water, such as in a waterfall or a rapidly flowing river, has kinetic energy. (credit “dam”: modification of work by "Pascal"/Flickr; credit “waterfall”: modification of work by Frank Gualtieri)

We associate potential energy not only with the matter's location (such as a child sitting on a tree branch), but also with the matter's structure. A spring on the ground has potential energy if it is compressed; so does a tautly pulled rubber band. The very existence of living cells relies heavily on structural potential energy. On a chemical level, the bonds that hold the molecules' atoms together have potential energy. Remember that anabolic cellular pathways require energy to synthesize complex molecules from simpler ones, and catabolic pathways release energy when complex

molecules break down. That certain chemical bonds' breakdown can release energy implies that those bonds have potential energy. In fact, there is potential energy stored within the bonds of all the food molecules we eat, which we eventually harness for use. This is because these bonds can release energy when broken. Scientists call the potential energy type that exists within chemical bonds that releases when those bonds break **chemical energy** ([link](#)). Chemical energy is responsible for providing living cells with energy from food. Breaking the molecular bonds within fuel molecules brings about the energy's release.



The molecules in gasoline contain chemical energy within the chemical bonds. This energy transforms into kinetic energy that allows a car to race on a racetrack. (credit “car”: modification of work by Russell Trow)

Note:**Link to Learning**

Visit this [site](#) and select “A simple pendulum” on the menu (under “Harmonic Motion”) to see the shifting kinetic (K) and potential energy (U) of a pendulum in motion.

Free Energy

After learning that chemical reactions release energy when energy-storing bonds break, an important next question is how do we quantify and express the chemical reactions with the associated energy? How can we compare the energy that releases from one reaction to that of another reaction? We use a measurement of **free energy** to quantitate these energy transfers. Scientists call this free energy Gibbs free energy (abbreviated with the letter G) after Josiah Willard Gibbs, the scientist who developed the measurement. Recall that according to the second law of thermodynamics, all energy transfers involve losing some energy in an unusable form such as heat, resulting in entropy. Gibbs free energy specifically refers to the energy that takes place with a chemical reaction that is available after we account for entropy. In other words, Gibbs free energy is usable energy, or energy that is available to do work.

Every chemical reaction involves a change in free energy, called delta G (ΔG). We can calculate the change in free energy for any system that undergoes such a change, such as a chemical reaction. To calculate ΔG , subtract the amount of energy lost to entropy (denoted as ΔS) from the system's total energy change. Scientists call this total energy change in the system **enthalpy** and we denote it as ΔH . The formula for calculating ΔG is as follows, where the symbol T refers to absolute temperature in Kelvin (degrees Celsius + 273):

Equation:

$$\Delta G = \Delta H - T\Delta S$$

We express a chemical reaction's standard free energy change as an amount of energy per mole of the reaction product (either in kilojoules or kilocalories, kJ/mol or kcal/mol; 1 kJ = 0.239 kcal) under standard pH, temperature, and pressure conditions. We generally calculate standard pH, temperature, and pressure conditions at pH 7.0 in biological systems, 25 degrees Celsius, and 100 kilopascals (1 atm pressure), respectively. Note that cellular conditions vary considerably from these standard conditions, and so standard calculated ΔG values for biological reactions will be different inside the cell.

Endergonic Reactions and Exergonic Reactions

If energy releases during a chemical reaction, then the resulting value from the above equation will be a negative number. In other words, reactions that release energy have a $\Delta G < 0$. A negative ΔG also means that the reaction's products have less free energy than the reactants, because they gave off some free energy during the reaction. Scientists call reactions that have a negative ΔG and consequently release free energy **exergonic reactions**. Think: *exergonic* means energy is *exiting* the system. We also refer to these reactions as spontaneous reactions, because they can occur without adding energy into the system. Understanding which chemical reactions are spontaneous and release free energy is extremely useful for biologists, because these reactions can be harnessed to perform work inside the cell. We must draw an important distinction between the term spontaneous and the idea of a chemical reaction that occurs immediately. Contrary to the everyday use of the term, a spontaneous reaction is not one that suddenly or quickly occurs. Rusting iron is an example of a spontaneous reaction that occurs slowly, little by little, over time.

If a chemical reaction requires an energy input rather than releasing energy, then the ΔG for that reaction will be a positive value. In this case, the products have more free energy than the reactants. Thus, we can think of the reactions' products as energy-storing molecules. We call these chemical reactions **endergonic reactions**, and they are non-spontaneous. An

endergonic reaction will not take place on its own without adding free energy.

Let's revisit the example of the synthesis and breakdown of the food molecule, glucose. Remember that building complex molecules, such as sugars, from simpler ones is an anabolic process and requires energy. Therefore, the chemical reactions involved in anabolic processes are endergonic reactions. Alternatively the catabolic process of breaking sugar down into simpler molecules releases energy in a series of exergonic reactions. Like the rust example above, the sugar breakdown involves spontaneous reactions, but these reactions do not occur instantaneously. [\[link\]](#) shows some other examples of endergonic and exergonic reactions. Later sections will provide more information about what else is required to make even spontaneous reactions happen more efficiently.

Note:

Visual Connection



(a)



(b)



(c)



(d)

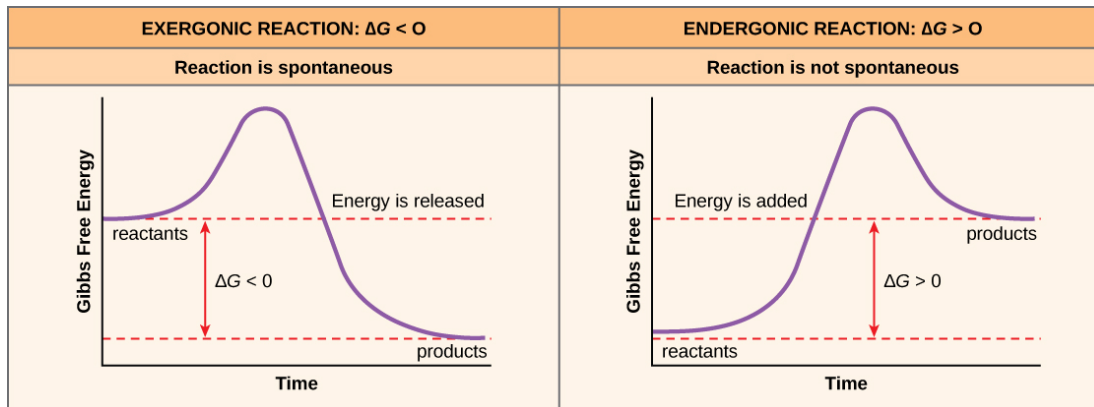
This figure shows some examples of endergonic processes (ones that require energy) and exergonic processes (ones that

release energy). These include (a) a compost pile decomposing, (b) a chick developing from a fertilized egg, (c) sand art destruction, and (d) a ball rolling down a hill. (credit a: modification of work by Natalie Maynor; credit b: modification of work by USDA; credit c: modification of work by “Athlex”/Flickr; credit d: modification of work by Harry Malsch)

Look at each of the processes, and decide if it is endergonic or exergonic. In each case, does enthalpy increase or decrease, and does entropy increase or decrease?

An important concept in studying metabolism and energy is that of chemical equilibrium. Most chemical reactions are reversible. They can proceed in both directions, releasing energy into their environment in one direction, and absorbing it from the environment in the other direction ([link](#)). The same is true for the chemical reactions involved in cell metabolism, such as the breaking down and building up of proteins into and from individual amino acids, respectively. Reactants within a closed system will undergo chemical reactions in both directions until they reach a state of equilibrium, which is one of the lowest possible free energy and a state of maximal entropy. To push the reactants and products away from a state of equilibrium requires energy. Either reactants or products must be added, removed, or changed. If a cell were a closed system, its chemical reactions would reach equilibrium, and it would die because there would be insufficient free energy left to perform the necessary work to maintain life. In a living cell, chemical reactions are constantly moving towards equilibrium, but never reach it. This is because a living cell is an open system. Materials pass in and out, the cell recycles the products of certain chemical reactions into other reactions, and there is never chemical equilibrium. In this way, living organisms are in a constant energy-requiring, uphill battle against equilibrium and entropy. This constant

energy supply ultimately comes from sunlight, which produces nutrients in the photosynthesis process.



Exergonic and endergonic reactions result in changes in Gibbs free energy. Exergonic reactions release energy. Endergonic reactions require energy to proceed.

Activation Energy

There is another important concept that we must consider regarding endergonic and exergonic reactions. Even exergonic reactions require a small amount of energy input before they can proceed with their energy-releasing steps. These reactions have a net release of energy, but still require some initial energy. Scientists call this small amount of energy input necessary for all chemical reactions to occur the **activation energy** (or free energy of activation) abbreviated as E_A ([link](#)).

Why would an energy-releasing, negative ΔG reaction actually require some energy to proceed? The reason lies in the steps that take place during a chemical reaction. During chemical reactions, certain chemical bonds break and new ones form. For example, when a glucose molecule breaks down, bonds between the molecule's carbon atoms break. Since these are energy-storing bonds, they release energy when broken. However, to get

them into a state that allows the bonds to break, the molecule must be somewhat contorted. A small energy input is required to achieve this contorted state. This contorted state is the **transition state**, and it is a high-energy, unstable state. For this reason, reactant molecules do not last long in their transition state, but very quickly proceed to the chemical reaction's next steps. Free energy diagrams illustrate the energy profiles for a given reaction. Whether the reaction is exergonic or endergonic determines whether the products in the diagram will exist at a lower or higher energy state than both the reactants and the products. However, regardless of this measure, the transition state of the reaction exists at a higher energy state than the reactants, and thus, E_A is always positive.

Note:

Link to Learning

Watch an animation of the move from free energy to transition state at [this site](#).

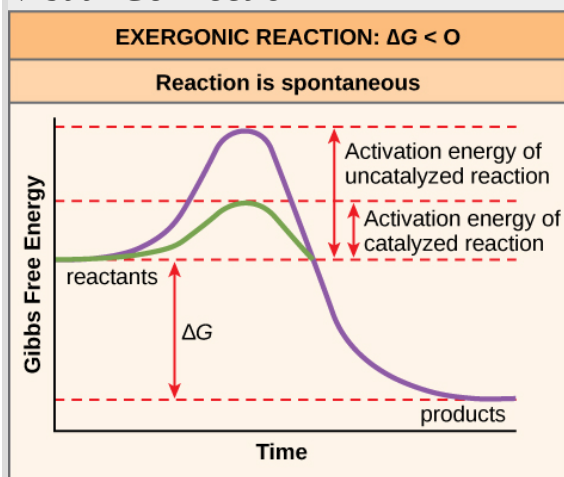
From where does the activation energy that chemical reactants require come? The activation energy's required source to push reactions forward is typically heat energy from the surroundings. **Heat energy** (the total bond energy of reactants or products in a chemical reaction) speeds up the molecule's motion, increasing the frequency and force with which they collide. It also moves atoms and bonds within the molecule slightly, helping them reach their transition state. For this reason, heating a system will cause chemical reactants within that system to react more frequently. Increasing the pressure on a system has the same effect. Once reactants have absorbed enough heat energy from their surroundings to reach the transition state, the reaction will proceed.

The activation energy of a particular reaction determines the rate at which it will proceed. The higher the activation energy, the slower the chemical reaction. The example of iron rusting illustrates an inherently slow reaction. This reaction occurs slowly over time because of its high E_A . Additionally, burning many fuels, which is strongly exergonic, will take place at a

negligible rate unless sufficient heat from a spark overcomes their activation energy. However, once they begin to burn, the chemical reactions release enough heat to continue the burning process, supplying the activation energy for surrounding fuel molecules. Like these reactions outside of cells, the activation energy for most cellular reactions is too high for heat energy to overcome at efficient rates. In other words, in order for important cellular reactions to occur at appreciable rates (number of reactions per unit time), their activation energies must be lowered ([link](#)). Scientist refer to this as catalysis. This is a very good thing as far as living cells are concerned. Important macromolecules, such as proteins, DNA, and RNA, store considerable energy, and their breakdown is exergonic. If cellular temperatures alone provided enough heat energy for these exergonic reactions to overcome their activation barriers, the cell's essential components would disintegrate.

Note:

Visual Connection



Activation energy is the energy required for a reaction to proceed, and it is lower if the reaction is catalyzed. This diagram's horizontal axis describes the sequence of events in time.

If no activation energy were required to break down sucrose (table sugar), would you be able to store it in a sugar bowl?

Section Summary

Energy comes in many different forms. Objects in motion do physical work, and kinetic energy is the energy of objects in motion. Objects that are not in motion may have the potential to do work, and thus, have potential energy. Molecules also have potential energy because breaking molecular bonds has the potential to release energy. Living cells depend on harvesting potential energy from molecular bonds to perform work. Free energy is a measure of energy that is available to do work. A system's free energy changes during energy transfers such as chemical reactions, and scientists refer to this change as ΔG .

A reaction's ΔG can be negative or positive, meaning that the reaction releases energy or consumes energy, respectively. A reaction with a negative ΔG that gives off energy is an exergonic reaction. One with a positive ΔG that requires energy input is an endergonic reaction. Exergonic reactions are spontaneous because their products have less energy than their reactants. Endergonic reactions' products have a higher energy state than the reactants, and so these are nonspontaneous reactions. However, all reactions (including spontaneous $-\Delta G$ reactions) require an initial energy input in order to reach the transition state, at which they will proceed. This initial input of energy is the activation energy.

Visual Connection Questions

Exercise:

Problem:

[\[link\]](#) Look at each of the processes, and decide if it is endergonic or exergonic. In each case, does enthalpy increase or decrease, and does entropy increase or decrease?

Solution:

[\[link\]](#) A compost pile decomposing is an exergonic process; enthalpy increases (energy is released) and entropy increases (large molecules are broken down into smaller ones). A baby developing from a fertilized egg is an endergonic process; enthalpy decreases (energy is absorbed) and entropy decreases. Sand art being destroyed is an exergonic process; there is no change in enthalpy, but entropy increases. A ball rolling downhill is an exergonic process; enthalpy decreases (energy is released), but there is no change in entropy.

Exercise:**Problem:**

[\[link\]](#) If no activation energy were required to break down sucrose (table sugar), would you be able to store it in a sugar bowl?

Solution:

[\[link\]](#) No. We can store chemical energy because of the need to overcome the barrier to its breakdown.

Review Questions**Exercise:**

Problem:

Consider a pendulum swinging. Which type(s) of energy is/are associated with the pendulum in the following instances: i. the moment at which it completes one cycle, just before it begins to fall back towards the other end, ii. the moment that it is in the middle between the two ends, and iii. just before it reaches the end of one cycle (just before instant i.).

- a. i. potential and kinetic, ii. potential and kinetic, iii. kinetic
- b. i. potential, ii. potential and kinetic, iii. potential and kinetic
- c. i. potential, ii. kinetic, iii. potential and kinetic
- d. i. potential and kinetic, ii. kinetic iii. kinetic

Solution:

C

Exercise:**Problem:**

Which of the following comparisons or contrasts between endergonic and exergonic reactions is false?

- a. Endergonic reactions have a positive ΔG and exergonic reactions have a negative ΔG .
- b. Endergonic reactions consume energy and exergonic reactions release energy.
- c. Both endergonic and exergonic reactions require a small amount of energy to overcome an activation barrier.
- d. Endergonic reactions take place slowly and exergonic reactions take place quickly.

Solution:

D

Exercise:**Problem:**

Which of the following is the best way to judge the relative activation energies between two given chemical reactions?

- a. Compare the ΔG values between the two reactions.
- b. Compare their reaction rates.
- c. Compare their ideal environmental conditions.
- d. Compare the spontaneity between the two reactions.

Solution:

B

Critical Thinking Questions**Exercise:****Problem:**

Explain in your own words the difference between a spontaneous reaction and one that occurs instantaneously, and what causes this difference.

Solution:

A spontaneous reaction is one that has a negative ΔG and thus releases energy. However, a spontaneous reaction need not occur quickly or suddenly like an instantaneous reaction. It may occur over long periods due to a large energy of activation, which prevents the reaction from occurring quickly.

Exercise:

Problem:

Describe the position of the transition state on a vertical energy scale, from low to high, relative to the position of the reactants and products, for both endergonic and exergonic reactions.

Solution:

The transition state is always higher in energy than the reactants and the products of a reaction (therefore, above), regardless of whether the reaction is endergonic or exergonic.

Glossary

activation energy

energy necessary for reactions to occur

chemical energy

potential energy in chemical bonds that releases when those bonds are broken

endergonic

describes chemical reactions that require energy input

enthalpy

a system's total energy

exergonic

describes chemical reactions that release free energy

free energy

Gibbs free energy is the usable energy, or energy that is available to do work

heat energy

total bond energy of reactants or products in a chemical reaction

kinetic energy

energy type that takes place with objects or particles in motion

potential energy

energy type that has the potential to do work; stored energy

transition state

high-energy, unstable state (an intermediate form between the substrate and the product) occurring during a chemical reaction

The Laws of Thermodynamics

By the end of this section, you will be able to do the following:

- Discuss the concept of entropy
- Explain the first and second laws of thermodynamics

Thermodynamics refers to the study of energy and energy transfer involving physical matter. The matter and its environment relevant to a particular case of energy transfer are classified as a system, and everything outside that system is the surroundings. For instance, when heating a pot of water on the stove, the system includes the stove, the pot, and the water. Energy transfers within the system (between the stove, pot, and water). There are two types of systems: open and closed. An open system is one in which energy can transfer between the system and its surroundings. The stovetop system is open because it can lose heat into the air. A closed system is one that cannot transfer energy to its surroundings.

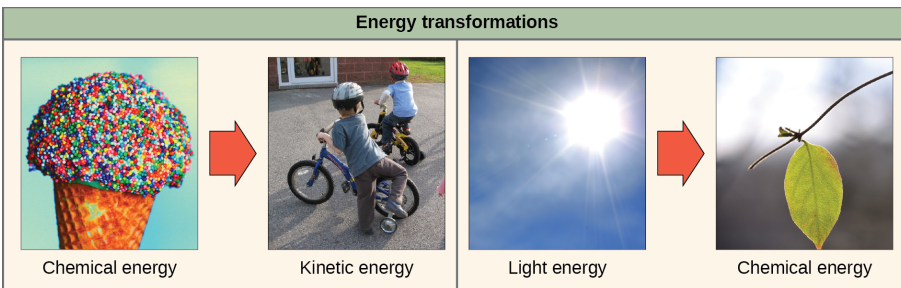
Biological organisms are open systems. Energy exchanges between them and their surroundings, as they consume energy-storing molecules and release energy to the environment by doing work. Like all things in the physical world, energy is subject to the laws of physics. The laws of thermodynamics govern the transfer of energy in and among all systems in the universe.

The First Law of Thermodynamics

The first law of thermodynamics deals with the total amount of energy in the universe. It states that this total amount of energy is constant. In other words, there has always been, and always will be, exactly the same amount of energy in the universe. Energy exists in many different forms. According to the first law of thermodynamics, energy may transfer from place to place or transform into different forms, but it cannot be created or destroyed. The transfers and transformations of energy take place around us all the time. Light bulbs transform electrical energy into light energy. Gas stoves transform chemical energy from natural gas into heat energy. Plants perform one of the most biologically useful energy transformations on earth: that of

converting sunlight energy into the chemical energy stored within organic molecules ([\[link\]](#)). [\[link\]](#) examples of energy transformations.

The challenge for all living organisms is to obtain energy from their surroundings in forms that they can transfer or transform into usable energy to do work. Living cells have evolved to meet this challenge very well. Chemical energy stored within organic molecules such as sugars and fats transforms through a series of cellular chemical reactions into energy within ATP molecules. Energy in ATP molecules is easily accessible to do work. Examples of the types of work that cells need to do include building complex molecules, transporting materials, powering the beating motion of cilia or flagella, contracting muscle fibers to create movement, and reproduction.



Here are two examples of energy transferring from one system to another and transformed from one form to another. Humans can convert the chemical energy in food, like this ice cream cone, into kinetic energy (the energy of movement to ride a bicycle). Plants can convert electromagnetic radiation (light energy) from the sun into chemical energy. (credit “ice cream”: modification of work by D. Sharon Pruitt; credit “kids on bikes”: modification of work by Michelle Riggen-Ransom; credit “leaf”: modification of work by Cory Zanker)

The Second Law of Thermodynamics

A living cell's primary tasks of obtaining, transforming, and using energy to do work may seem simple. However, the second law of thermodynamics explains why these tasks are harder than they appear. None of the energy transfers that we have discussed, along with all energy transfers and transformations in the universe, is completely efficient. In every energy transfer, some amount of energy is lost in a form that is unusable. In most cases, this form is heat energy. Thermodynamically, scientists define **heat energy** as energy that transfers from one system to another that is not doing work. For example, when an airplane flies through the air, it loses some of its energy as heat energy due to friction with the surrounding air. This friction actually heats the air by temporarily increasing air molecule speed. Likewise, some energy is lost as heat energy during cellular metabolic reactions. This is good for warm-blooded creatures like us, because heat energy helps to maintain our body temperature. Strictly speaking, no energy transfer is completely efficient, because some energy is lost in an unusable form.

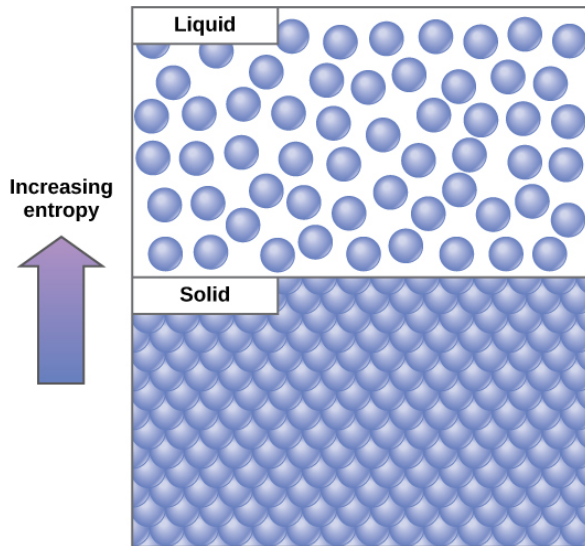
An important concept in physical systems is that of order and disorder (or randomness). The more energy that a system loses to its surroundings, the less ordered and more random the system. Scientists refer to the measure of randomness or disorder within a system as **entropy**. High entropy means high disorder and low energy ([\[link\]](#)). To better understand entropy, think of a student's bedroom. If no energy or work were put into it, the room would quickly become messy. It would exist in a very disordered state, one of high entropy. Energy must be put into the system, in the form of the student doing work and putting everything away, in order to bring the room back to a state of cleanliness and order. This state is one of low entropy. Similarly, a car or house must be constantly maintained with work in order to keep it in an ordered state. Left alone, a house's or car's entropy gradually increases through rust and degradation. Molecules and chemical reactions have varying amounts of entropy as well. For example, as chemical reactions reach a state of equilibrium, entropy increases, and as molecules at a high concentration in one place diffuse and spread out, entropy also increases.

Note:**Scientific Connection****Transfer of Energy and the Resulting Entropy**

Set up a simple experiment to understand how energy transfers and how a change in entropy results.

1. Take a block of ice. This is water in solid form, so it has a high structural order. This means that the molecules cannot move very much and are in a fixed position. The ice's temperature is 0°C . As a result, the system's entropy is low.
2. Allow the ice to melt at room temperature. What is the state of molecules in the liquid water now? How did the energy transfer take place? Is the system's entropy higher or lower? Why?
3. Heat the water to its boiling point. What happens to the system's entropy when the water is heated?

Think of all physical systems of in this way: Living things are highly ordered, requiring constant energy input to maintain themselves in a state of low entropy. As living systems take in energy-storing molecules and transform them through chemical reactions, they lose some amount of usable energy in the process, because no reaction is completely efficient. They also produce waste and by-products that are not useful energy sources. This process increases the entropy of the system's surroundings. Since all energy transfers result in losing some usable energy, the second law of thermodynamics states that every energy transfer or transformation increases the universe's entropy. Even though living things are highly ordered and maintain a state of low entropy, the universe's entropy in total is constantly increasing due to losing usable energy with each energy transfer that occurs. Essentially, living things are in a continuous uphill battle against this constant increase in universal entropy.



Entropy is a measure of randomness or disorder in a system. Gases have higher entropy than liquids, and liquids have higher entropy than solids.

Section Summary

In studying energy, scientists use the term “system” to refer to the matter and its environment involved in energy transfers. Everything outside of the system is the surroundings. Single cells are biological systems. We can think of systems as having a certain amount of order. It takes energy to make a system more ordered. The more ordered a system, the lower its entropy. Entropy is a measure of a system's disorder. As a system becomes more disordered, the lower its energy and the higher its entropy.

The laws of thermodynamics are a series of laws that describe the properties and processes of energy transfer. The first law states that the total amount of energy in the universe is constant. This means that energy cannot be created or destroyed, only transferred or transformed. The second law of thermodynamics states that every energy transfer involves some loss of energy in an unusable form, such as heat energy, resulting in a more

disordered system. In other words, no energy transfer is completely efficient, and all transfers trend toward disorder.

Review Questions

Exercise:

Problem:

Which of the following is not an example of an energy transformation?

- a. turning on a light switch
- b. solar panels at work
- c. formation of static electricity
- d. none of the above

Solution:

A

Exercise:

Problem:

In each of the three systems, determine the state of entropy (low or high) when comparing the first and second: i. the instant that a perfume bottle is sprayed compared with 30 seconds later, ii. an old 1950s car compared with a brand new car, and iii. a living cell compared with a dead cell.

- a. i. low, ii. high, iii. low
- b. i. low, ii. high, iii. high
- c. i. high, ii. low, iii. high
- d. i. high, ii. low, iii. low

Solution:

A

Critical Thinking Questions

Exercise:

Problem:

Imagine an elaborate ant farm with tunnels and passageways through the sand where ants live in a large community. Now imagine that an earthquake shook the ground and demolished the ant farm. In which of these two scenarios, before or after the earthquake, was the ant farm system in a state of higher or lower entropy?

Solution:

The ant farm had lower entropy before the earthquake because it was a highly ordered system. After the earthquake, the system became much more disordered and had higher entropy.

Exercise:

Problem:

Energy transfers take place constantly in everyday activities. Think of two scenarios: cooking on a stove and driving. Explain how the second law of thermodynamics applies to these two scenarios.

Solution:

While cooking, food is heating up on the stove, but not all of the heat goes to cooking the food, some of it is lost as heat energy to the surrounding air, increasing entropy. While driving, cars burn gasoline to run the engine and move the car. This reaction is not completely efficient, as some energy during this process is lost as heat energy, which is why the hood and the components underneath it heat up while the engine is turned on. The tires also heat up because of friction with the pavement, which is additional energy loss. This energy transfer, like all others, also increases entropy.

Glossary

entropy (S)

measure of randomness or disorder within a system

heat

energy transferred from one system to another that is not work (energy of the molecules' motion or particles)

thermodynamics

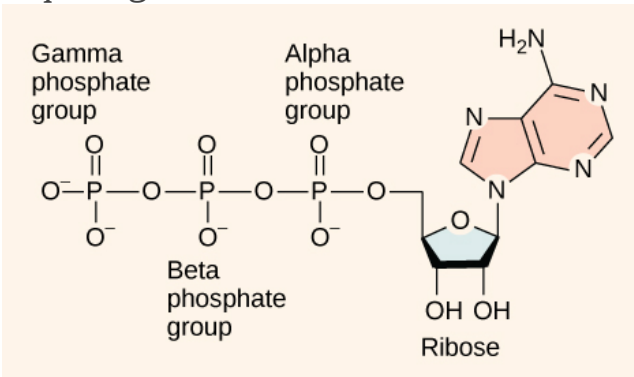
study of energy and energy transfer involving physical matter

ATP: Adenosine Triphosphate

By the end of this section, you will be able to do the following:

- Explain ATP's role as the cellular energy currency
- Describe how energy releases through ATP hydrolysis

Even exergonic, energy-releasing reactions require a small amount of activation energy in order to proceed. However, consider endergonic reactions, which require much more energy input, because their products have more free energy than their reactants. Within the cell, from where does energy to power such reactions come? The answer lies with an energy-supplying molecule scientists call **adenosine triphosphate**, or **ATP**. This is a small, relatively simple molecule ([\[link\]](#)), but within some of its bonds, it contains the potential for a quick burst of energy that can be harnessed to perform cellular work. Think of this molecule as the cells' primary energy currency in much the same way that money is the currency that people exchange for things they need. ATP powers the majority of energy-requiring cellular reactions.



ATP is the cell's primary energy currency. It has an adenosine backbone with three phosphate groups attached.

As its name suggests, adenosine triphosphate is comprised of adenosine bound to three phosphate groups ([\[link\]](#)). Adenosine is a nucleoside consisting of the nitrogenous base adenine and a five-carbon sugar, ribose.

The three phosphate groups, in order of closest to furthest from the ribose sugar, are alpha, beta, and gamma. Together, these chemical groups constitute an energy powerhouse. However, not all bonds within this molecule exist in a particularly high-energy state. Both bonds that link the phosphates are equally high-energy bonds (**phosphoanhydride bonds**) that, when broken, release sufficient energy to power a variety of cellular reactions and processes. These high-energy bonds are the bonds between the second and third (or beta and gamma) phosphate groups and between the first and second phosphate groups. These bonds are “high-energy” because the products of such bond breaking—adenosine diphosphate (ADP) and one inorganic phosphate group (P_i)—have considerably lower free energy than the reactants: ATP and a water molecule. Because this reaction takes place using a water molecule, it is a hydrolysis reaction. In other words, ATP hydrolyzes into ADP in the following reaction:

Equation:



Like most chemical reactions, ATP to ADP hydrolysis is reversible. The reverse reaction regenerates ATP from $\text{ADP} + P_i$. Cells rely on ATP regeneration just as people rely on regenerating spent money through some sort of income. Since ATP hydrolysis releases energy, ATP regeneration must require an input of free energy. This equation expresses ATP formation:

Equation:

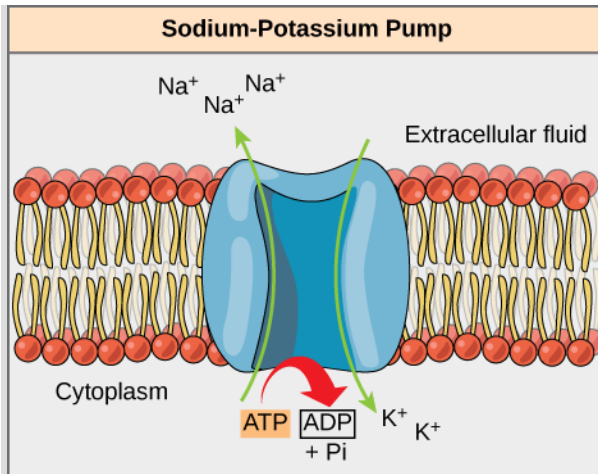


Two prominent questions remain with regard to using ATP as an energy source. Exactly how much free energy releases with ATP hydrolysis, and how does that free energy do cellular work? The calculated ΔG for the hydrolysis of one ATP mole into ADP and P_i is -7.3 kcal/mole (-30.5 kJ/mol). Since this calculation is true under standard conditions, one would expect a different value exists under cellular conditions. In fact, the ΔG for one ATP mole's hydrolysis in a living cell is almost double the value at standard conditions: -14 kcal/mol (-57 kJ/mol).

ATP is a highly unstable molecule. Unless quickly used to perform work, ATP spontaneously dissociates into ADP + P_i , and the free energy released during this process is lost as heat. The second question we posed above discusses how ATP hydrolysis energy release performs work inside the cell. This depends on a strategy scientists call energy coupling. Cells couple the ATP hydrolysis' exergonic reaction allowing them to proceed. One example of energy coupling using ATP involves a transmembrane ion pump that is extremely important for cellular function. This sodium-potassium pump (Na^+/K^+ pump) drives sodium out of the cell and potassium into the cell ([link](#)). A large percentage of a cell's ATP powers this pump, because cellular processes bring considerable sodium into the cell and potassium out of it. The pump works constantly to stabilize cellular concentrations of sodium and potassium. In order for the pump to turn one cycle (exporting three Na^+ ions and importing two K^+ ions), one ATP molecule must hydrolyze. When ATP hydrolyzes, its gamma phosphate does not simply float away, but it actually transfers onto the pump protein. Scientists call this process of a phosphate group binding to a molecule phosphorylation. As with most ATP hydrolysis cases, a phosphate from ATP transfers onto another molecule. In a phosphorylated state, the Na^+/K^+ pump has more free energy and is triggered to undergo a conformational change. This change allows it to release Na^+ to the cell's outside. It then binds extracellular K^+ , which, through another conformational change, causes the phosphate to detach from the pump. This phosphate release triggers the K^+ to release to the cell's inside. Essentially, the energy released from the ATP hydrolysis couples with the energy required to power the pump and transport Na^+ and K^+ ions. ATP performs cellular work using this basic form of energy coupling through phosphorylation.

Note:

Visual Connection



The sodium-potassium pump is an example of energy coupling.

The energy derived from exergonic ATP hydrolysis pumps sodium and potassium ions across the cell membrane.

One ATP molecule's hydrolysis releases 7.3 kcal/mol of energy ($\Delta G = -7.3$ kcal/mol of energy). If it takes 2.1 kcal/mol of energy to move one Na^+ across the membrane ($\Delta G = +2.1$ kcal/mol of energy), how many sodium ions could one ATP molecule's hydrolysis move?

Often during cellular metabolic reactions, such as nutrient synthesis and breakdown, certain molecules must alter slightly in their conformation to become substrates for the next step in the reaction series. One example is during the very first steps of cellular respiration, when a sugar glucose molecule breaks down in the process of glycolysis. In the first step, ATP is required to phosphorylate glucose, creating a high-energy but unstable intermediate. This phosphorylation reaction powers a conformational change that allows the phosphorylated glucose molecule to convert to the phosphorylated sugar fructose. Fructose is a necessary intermediate for glycolysis to move forward. Here, ATP hydrolysis' exergonic reaction couples with the endergonic reaction of converting glucose into a

phosphorylated intermediate in the pathway. Once again, the energy released by breaking a phosphate bond within ATP was used for phosphorylating another molecule, creating an unstable intermediate and powering an important conformational change.

Note:**Link to Learning**

See an interactive animation of the ATP-producing glycolysis process at this [site](#).

Section Summary

ATP is the primary energy-supplying molecule for living cells. ATP is comprised of a nucleotide, a five-carbon sugar, and three phosphate groups. The bonds that connect the phosphates (phosphoanhydride bonds) have high-energy content. The energy released from ATP hydrolysis into ADP + P_i performs cellular work. Cells use ATP to perform work by coupling ATP hydrolysis' exergonic reaction with endergonic reactions. ATP donates its phosphate group to another molecule via phosphorylation. The phosphorylated molecule is at a higher-energy state and is less stable than its unphosphorylated form, and this added energy from phosphate allows the molecule to undergo its endergonic reaction.

Visual Connection Questions

Exercise:**Problem:**

[\[link\]](#) One ATP molecule's hydrolysis releases 7.3 kcal/mol of energy ($\Delta G = -7.3$ kcal/mol of energy). If it takes 2.1 kcal/mol of energy to move one Na^+ across the membrane ($\Delta G = +2.1$ kcal/mol of energy), how many sodium ions could one ATP molecule's hydrolysis move?

Solution:

[\[link\]](#) Three sodium ions could be moved by the hydrolysis of one ATP molecule. The ΔG of the coupled reaction must be negative.

Movement of three sodium ions across the membrane will take 6.3 kcal of energy ($2.1 \text{ kcal} \times 3 \text{ Na}^+ \text{ ions} = 6.3 \text{ kcal}$). Hydrolysis of ATP provides 7.3 kcal of energy, more than enough to power this reaction. Movement of four sodium ions across the membrane, however, would require 8.4 kcal of energy, more than one ATP molecule can provide.

Review Questions

Exercise:

Problem: The energy released by the hydrolysis of ATP is_____

- a. primarily stored between the alpha and beta phosphates
- b. equal to -57 kcal/mol
- c. harnessed as heat energy by the cell to perform work
- d. providing energy to coupled reactions

Solution:

D

Exercise:**Problem:**

Which of the following molecules is likely to have the most potential energy?

- a. sucrose
- b. ATP
- c. glucose
- d. ADP

Solution:

A

Critical Thinking Questions

Exercise:**Problem:**

Do you think that the E_A for ATP hydrolysis is relatively low or high? Explain your reasoning.

Solution:

The activation energy for hydrolysis is very low. Not only is ATP hydrolysis an exergonic process with a large $-\Delta G$, but ATP is also a very unstable molecule that rapidly breaks down into ADP + P_i if not utilized quickly. This suggests a very low E_A since it hydrolyzes so quickly.

Glossary

ATP

adenosine triphosphate, the cell's energy currency

phosphoanhydride bond

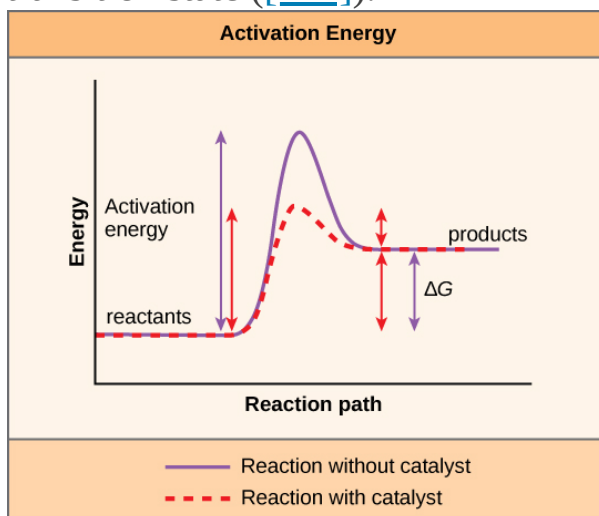
bond that connects phosphates in an ATP molecule

Enzymes

By the end of this section, you will be able to do the following:

- Describe the role of enzymes in metabolic pathways
- Explain how enzymes function as molecular catalysts
- Discuss enzyme regulation by various factors

A substance that helps a chemical reaction to occur is a catalyst, and the special molecules that catalyze biochemical reactions are enzymes. Almost all enzymes are proteins, comprised of amino acid chains, and they perform the critical task of lowering the activation energies of chemical reactions inside the cell. Enzymes do this by binding to the reactant molecules, and holding them in such a way as to make the chemical bond-breaking and bond-forming processes take place more readily. It is important to remember that enzymes do not change the reaction's ΔG . In other words, they do not change whether a reaction is exergonic (spontaneous) or endergonic. This is because they do not change the reactants' or products' free energy. They only reduce the activation energy required to reach the transition state ([link](#)).



Enzymes lower the reaction's activation energy but do not change the reaction's free energy.

Enzyme Active Site and Substrate Specificity

The chemical reactants to which an enzyme binds are the enzyme's **substrates**. There may be one or more substrates, depending on the particular chemical reaction. In some reactions, a single-reactant substrate breaks down into multiple products. In others, two substrates may come together to create one larger molecule. Two reactants might also enter a reaction, both become modified, and leave the reaction as two products. The location within the enzyme where the substrate binds is the enzyme's **active site**. This is where the "action" happens. Since enzymes are proteins, there is a unique combination of amino acid residues (also side chains, or R groups) within the active site. Different properties characterize each residue. These can be large or small, weakly acidic or basic, hydrophilic or hydrophobic, positively or negatively charged, or neutral. The unique combination of amino acid residues, their positions, sequences, structures, and properties, creates a very specific chemical environment within the active site. This specific environment is suited to bind, albeit briefly, to a specific chemical substrate (or substrates). Due to this jigsaw puzzle-like match between an enzyme and its substrates (which adapts to find the best fit between the transition state and the active site), enzymes are known for their specificity. The "best fit" results from the shape and the amino acid functional group's attraction to the substrate. There is a specifically matched enzyme for each substrate and, thus, for each chemical reaction; however, there is flexibility as well.

The fact that active sites are so perfectly suited to provide specific environmental conditions also means that they are subject to local environmental influences. It is true that increasing the environmental temperature generally increases reaction rates, enzyme-catalyzed or otherwise. However, increasing or decreasing the temperature outside of an optimal range can affect chemical bonds within the active site in such a way that they are less well suited to bind substrates. High temperatures will eventually cause enzymes, like other biological molecules, to **denature**, a process that changes the substance's natural properties. Likewise, the local environment's pH can also affect enzyme function. Active site amino acid residues have their own acidic or basic properties that are optimal for catalysis. These residues are sensitive to changes in pH that can impair the

way substrate molecules bind. Enzymes are suited to function best within a certain pH range, and, as with temperature, extreme environmental pH values (acidic or basic) can cause enzymes to denature.

Induced Fit and Enzyme Function

For many years, scientists thought that enzyme-substrate binding took place in a simple “lock-and-key” fashion. This model asserted that the enzyme and substrate fit together perfectly in one instantaneous step. However, current research supports a more refined view scientists call **induced fit** ([link](#)). This model expands upon the lock-and-key model by describing a more dynamic interaction between enzyme and substrate. As the enzyme and substrate come together, their interaction causes a mild shift in the enzyme’s structure that confirms an ideal binding arrangement between the enzyme and the substrate's transition state. This ideal binding maximizes the enzyme’s ability to catalyze its reaction.

Note:

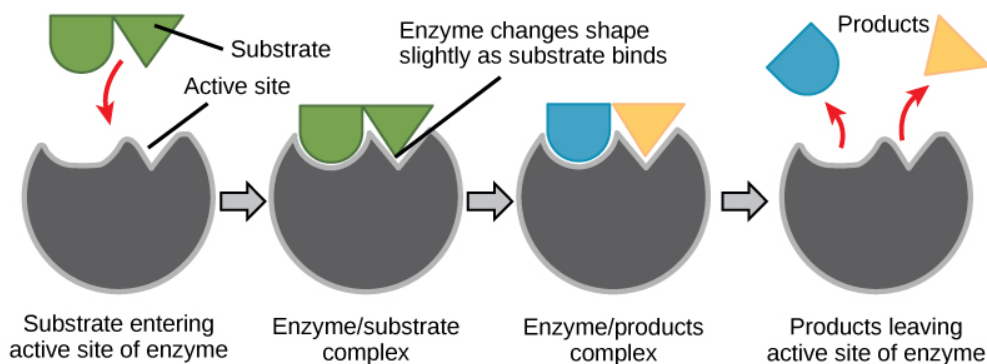
Link to Learning

View an induced fit animation at [this website](#).

When an enzyme binds its substrate, it forms an enzyme-substrate complex. This complex lowers the reaction's activation energy and promotes its rapid progression in one of many ways. On a basic level, enzymes promote chemical reactions that involve more than one substrate by bringing the substrates together in an optimal orientation. The appropriate region (atoms and bonds) of one molecule is juxtaposed to the other molecule's appropriate region with which it must react. Another way in which enzymes promote substrate reaction is by creating an optimal environment within the active site for the reaction to occur. Certain chemical reactions might proceed best in a slightly acidic or non-polar environment. The chemical properties that emerge from the particular arrangement of amino acid

residues within an active site create the perfect environment for an enzyme's specific substrates to react.

You have learned that the activation energy required for many reactions includes the energy involved in manipulating or slightly contorting chemical bonds so that they can easily break and allow others to reform. Enzymatic action can aid this process. The enzyme-substrate complex can lower the activation energy by contorting substrate molecules in such a way as to facilitate bond-breaking, helping to reach the transition state. Finally, enzymes can also lower activation energies by taking part in the chemical reaction itself. The amino acid residues can provide certain ions or chemical groups that actually form covalent bonds with substrate molecules as a necessary step of the reaction process. In these cases, it is important to remember that the enzyme will always return to its original state at the reaction's completion. One of enzymes' hallmark properties is that they remain ultimately unchanged by the reactions they catalyze. After an enzyme catalyzes a reaction, it releases its product(s).



According to the induced-fit model, both enzyme and substrate undergo dynamic conformational changes upon binding. The enzyme contorts the substrate into its transition state, thereby increasing the reaction's rate.

Metabolism Control Through Enzyme Regulation

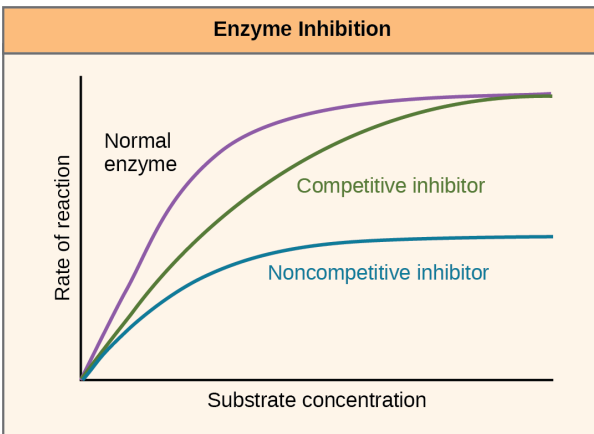
It would seem ideal to have a scenario in which all the encoded enzymes in an organism's genome existed in abundant supply and functioned optimally under all cellular conditions, in all cells, at all times. In reality, this is far from the case. A variety of mechanisms ensure that this does not happen. Cellular needs and conditions vary from cell to cell, and change within individual cells over time. The required enzymes and energetic demands of stomach cells are different from those of fat storage cells, skin cells, blood cells, and nerve cells. Furthermore, a digestive cell works much harder to process and break down nutrients during the time that closely follows a meal compared with many hours after a meal. As these cellular demands and conditions vary, so do the amounts and functionality of different enzymes.

Since the rates of biochemical reactions are controlled by activation energy, and enzymes lower and determine activation energies for chemical reactions, the relative amounts and functioning of the variety of enzymes within a cell ultimately determine which reactions will proceed and at which rates. This determination is tightly controlled. In certain cellular environments, environmental factors like pH and temperature partly control enzyme activity. There are other mechanisms through which cells control enzyme activity and determine the rates at which various biochemical reactions will occur.

Molecular Regulation of Enzymes

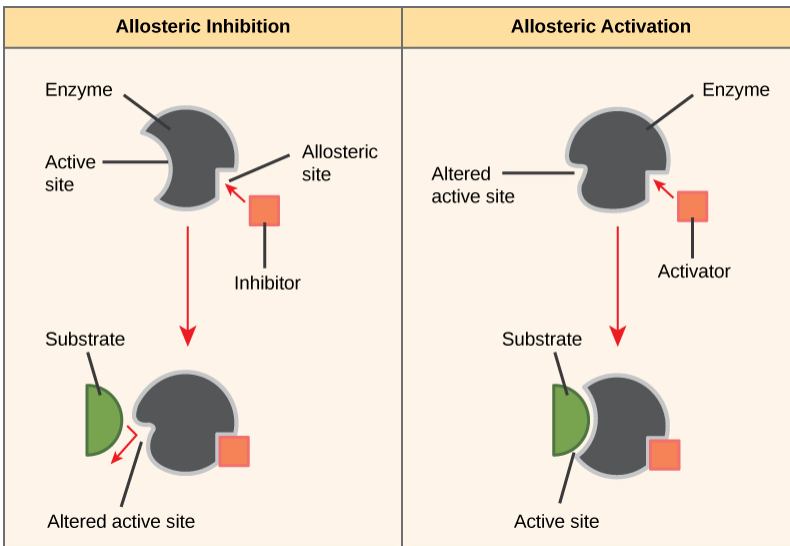
Enzymes can be regulated in ways that either promote or reduce their activity. There are many different kinds of molecules that inhibit or promote enzyme function, and various mechanisms exist for doing so. For example, in some cases of enzyme inhibition, an inhibitor molecule is similar enough to a substrate that it can bind to the active site and simply block the substrate from binding. When this happens, the enzyme is inhibited through **competitive inhibition**, because an inhibitor molecule competes with the substrate for active site binding ([\[link\]](#)). Alternatively, in noncompetitive inhibition, an inhibitor molecule binds to the enzyme in a location other

than an allosteric site, a binding site away from the active site, and still manages to block substrate binding to the active site.



Competitive and noncompetitive inhibition affect the reaction's rate differently. Competitive inhibitors affect the initial rate but do not affect the maximal rate; whereas, noncompetitive inhibitors affect the maximal rate.

Some inhibitor molecules bind to enzymes in a location where their binding induces a conformational change that reduces the enzyme's affinity for its substrate. This type of inhibition is an **allosteric inhibition** ([\[link\]](#)). More than one polypeptide comprise most allosterically regulated enzymes, meaning that they have more than one protein subunit. When an allosteric inhibitor binds to an enzyme, all active sites on the protein subunits change slightly such that they bind their substrates with less efficiency. There are allosteric activators as well as inhibitors. Allosteric activators bind to locations on an enzyme away from the active site, inducing a conformational change that increases the affinity of the enzyme's active site(s) for its substrate(s).



Allosteric inhibitors modify the enzyme's active site so that substrate binding is reduced or prevented. In contrast, allosteric activators modify the enzyme's active site so that the affinity for the substrate increases.

Note:
Everyday Connection



Have you ever wondered how

pharmaceutical drugs are developed? (credit: Deborah Austin)

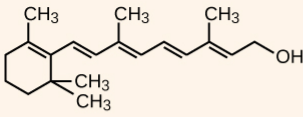
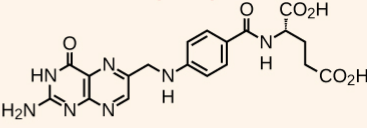
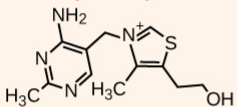
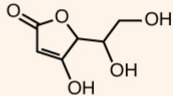
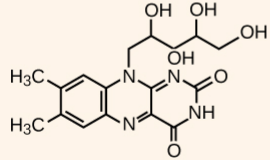
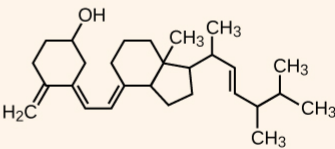
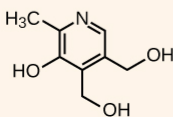
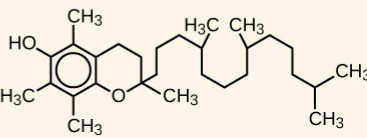
Drug Discovery by Looking for Inhibitors of Key Enzymes in Specific Pathways

Enzymes are key components of metabolic pathways. Understanding how enzymes work and how they can be regulated is a key principle behind developing many pharmaceutical drugs ([\[link\]](#)) on the market today. Biologists working in this field collaborate with other scientists, usually chemists, to design drugs.

Consider statins for example—which is a class of drugs that reduces cholesterol levels. These compounds are essentially inhibitors of the enzyme HMG-CoA reductase. HMG-CoA reductase is the enzyme that synthesizes cholesterol from lipids in the body. By inhibiting this enzyme, the drug reduces cholesterol levels synthesized in the body. Similarly, acetaminophen, popularly marketed under the brand name Tylenol, is an inhibitor of the enzyme cyclooxygenase. While it is effective in providing relief from fever and inflammation (pain), scientists still do not completely understand its mechanism of action.

How are drugs developed? One of the first challenges in drug development is identifying the specific molecule that the drug is intended to target. In the case of statins, HMG-CoA reductase is the drug target. Researchers identify targets through painstaking research in the laboratory. Identifying the target alone is not sufficient. Scientists also need to know how the target acts inside the cell and which reactions go awry in the case of disease. Once researchers identify the target and the pathway, then the actual drug design process begins. During this stage, chemists and biologists work together to design and synthesize molecules that can either block or activate a particular reaction. However, this is only the beginning: both if and when a drug prototype is successful in performing its function, then it must undergo many tests from *in vitro* experiments to clinical trials before it can obtain FDA approval to be on the market.

Many enzymes don't work optimally, or even at all, unless bound to other specific non-protein helper molecules, either temporarily through ionic or hydrogen bonds or permanently through stronger covalent bonds. Two types of helper molecules are **cofactors** and **coenzymes**. Binding to these molecules promotes optimal conformation and function for their respective enzymes. Cofactors are inorganic ions such as iron (Fe^{++}) and magnesium (Mg^{++}). One example of an enzyme that requires a metal ion as a cofactor is the enzyme that builds DNA molecules, DNA polymerase, which requires a bound zinc ion (Zn^{++}) to function. Coenzymes are organic helper molecules, with a basic atomic structure comprised of carbon and hydrogen, which are required for enzyme action. The most common sources of coenzymes are dietary vitamins ([\[link\]](#)). Some vitamins are precursors to coenzymes and others act directly as coenzymes. Vitamin C is a coenzyme for multiple enzymes that take part in building the important connective tissue component, collagen. An important step in breaking down glucose to yield energy is catalysis by a multi-enzyme complex scientists call pyruvate dehydrogenase. Pyruvate dehydrogenase is a complex of several enzymes that actually requires one cofactor (a magnesium ion) and five different organic coenzymes to catalyze its specific chemical reaction. Therefore, enzyme function is, in part, regulated by an abundance of various cofactors and coenzymes, which the diets of most organisms supply.

Dietary Vitamins	
Vitamin A (retinol) 	Folic acid (folate) 
Vitamin B₁ (thiamin) 	Vitamin C (ascorbic acid) 
Vitamin B₂ (riboflavin) 	Vitamin D₂ (calciferol) 
Vitamin B₆ (pyridoxine) 	Vitamin E (α-tocopherol) 

Vitamins are important coenzymes or precursors of coenzymes, and are required for enzymes to function properly. Multivitamin capsules usually contain mixtures of all the vitamins at different percentages.

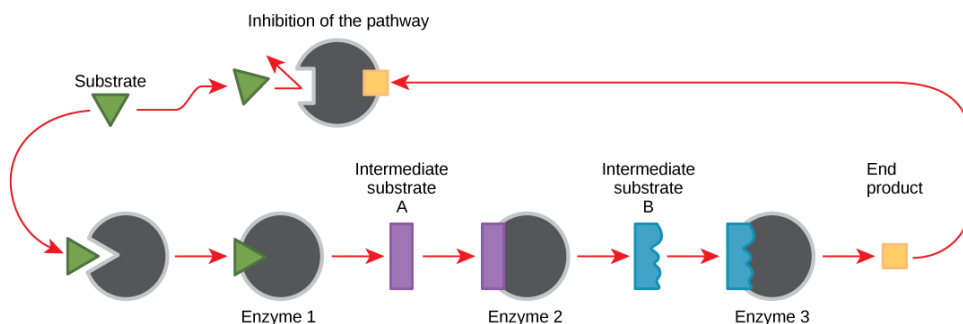
Enzyme Compartmentalization

In eukaryotic cells, molecules such as enzymes are usually compartmentalized into different organelles. This allows for yet another level of regulation of enzyme activity. Enzymes required only for certain cellular processes are sometimes housed separately along with their substrates, allowing for more efficient chemical reactions. Examples of this

sort of enzyme regulation based on location and proximity include the enzymes involved in the latter stages of cellular respiration, which take place exclusively in the mitochondria, and the enzymes involved in digesting cellular debris and foreign materials, located within lysosomes.

Feedback Inhibition in Metabolic Pathways

Molecules can regulate enzyme function in many ways. However, a major question remains: What are these molecules and from where do they come? Some are cofactors and coenzymes, ions, and organic molecules, as you have learned. What other molecules in the cell provide enzymatic regulation, such as allosteric modulation, and competitive and noncompetitive inhibition? The answer is that a wide variety of molecules can perform these roles. Some include pharmaceutical and non-pharmaceutical drugs, toxins, and poisons from the environment. Perhaps the most relevant sources of enzyme regulatory molecules, with respect to cellular metabolism, are cellular metabolic reaction products themselves. In a most efficient and elegant way, cells have evolved to use their own reactions' products for feedback inhibition of enzyme activity. **Feedback inhibition** involves using a reaction product to regulate its own further production ([\[link\]](#)). The cell responds to the abundance of specific products by slowing down production during anabolic or catabolic reactions. Such reaction products may inhibit the enzymes that catalyzed their production through the mechanisms that we described above.



Metabolic pathways are a series of reactions that

multiple enzymes catalyze. Feedback inhibition, where the pathway's end product inhibits an upstream step, is an important regulatory mechanism in cells.

Producing both amino acids and nucleotides is controlled through feedback inhibition. Additionally, ATP is an allosteric regulator of some of the enzymes involved in sugar's catabolic breakdown, the process that produces ATP. In this way, when ATP is abundant, the cell can prevent its further production. Remember that ATP is an unstable molecule that can spontaneously dissociate into ADP. If too much ATP were present in a cell, much of it would go to waste. Alternatively, ADP serves as a positive allosteric regulator (an allosteric activator) for some of the same enzymes that ATP inhibits. Thus, when relative ADP levels are high compared to ATP, the cell is triggered to produce more ATP through sugar catabolism.

Section Summary

Enzymes are chemical catalysts that accelerate chemical reactions at physiological temperatures by lowering their activation energy. Enzymes are usually proteins consisting of one or more polypeptide chains. Enzymes have an active site that provides a unique chemical environment, comprised of certain amino acid R groups (residues). This unique environment is perfectly suited to convert particular chemical reactants for that enzyme, scientists call substrates, into unstable intermediates that they call transition states. Enzymes and substrates bind with an induced fit, which means that enzymes undergo slight conformational adjustments upon substrate contact, leading to full, optimal binding. Enzymes bind to substrates and catalyze reactions in four different ways: bringing substrates together in an optimal orientation, compromising the bond structures of substrates so that bonds can break down more easily, providing optimal environmental conditions for a reaction to occur, or participating directly in their chemical reaction by forming transient covalent bonds with the substrates.

Enzyme action must be regulated so that in a given cell at a given time, the desired reactions catalyze and the undesired reactions are not. Enzymes are

regulated by cellular conditions, such as temperature and pH. They are also regulated through their location within a cell, sometimes compartmentalized so that they can only catalyze reactions under certain circumstances.

Enzyme inhibition and activation via other molecules are other important ways that enzymes are regulated. Inhibitors can act competitively, noncompetitively, or allosterically. Noncompetitive inhibitors are usually allosteric. Activators can also enhance enzyme function allosterically. The most common method by which cells regulate the enzymes in metabolic pathways is through feedback inhibition. During feedback inhibition, metabolic pathway products serve as inhibitors (usually allosteric) of one or more of the enzymes (usually the first committed enzyme of the pathway) involved in the pathway that produces them.

Review Questions

Exercise:

Problem: Which of the following is not true about enzymes:

- a. They increase ΔG of reactions.
- b. They are usually made of amino acids.
- c. They lower the activation energy of chemical reactions.
- d. Each one is specific to the particular substrate(s) to which it binds.

Solution:

A

Exercise:

Problem: An allosteric inhibitor does which of the following?

- a. Binds to an enzyme away from the active site and changes the conformation of the active site, increasing its affinity for substrate binding.

- b. Binds to the active site and blocks it from binding substrate.
- c. Binds to an enzyme away from the active site and changes the conformation of the active site, decreasing its affinity for the substrate.
- d. Binds directly to the active site and mimics the substrate.

Solution:

C

Exercise:

Problem:

Which of the following analogies best describes the induced-fit model of enzyme-substrate binding?

- a. a hug between two people
- b. a key fitting into a lock
- c. a square peg fitting through the square hole and a round peg fitting through the round hole of a children's toy
- d. the fitting together of two jigsaw puzzle pieces

Solution:

A

Critical Thinking Questions

Exercise:

Problem:

With regard to enzymes, why are vitamins necessary for good health? Give examples.

Solution:

Most vitamins and minerals act as coenzymes and cofactors for enzyme action. Many enzymes require the binding of certain cofactors or coenzymes to be able to catalyze their reactions. Since enzymes catalyze many important reactions, it is critical to obtain sufficient vitamins and minerals from the diet and from supplements. Vitamin C (ascorbic acid) is a coenzyme necessary for the action of enzymes that build collagen, an important protein component of connective tissue throughout the body. Magnesium ion (Mg^{++}) is an important cofactor that is necessary for the enzyme pyruvate dehydrogenase to catalyze part of the pathway that breaks down sugar to produce energy. Vitamins cannot be produced in the human body and therefore must be obtained in the diet.

Exercise:**Problem:**

Explain in your own words how enzyme feedback inhibition benefits a cell.

Solution:

Feedback inhibition allows cells to control the amounts of metabolic products produced. If there is too much of a particular product relative to the cell's needs, feedback inhibition effectively causes the cell to decrease production of that particular product. In general, this reduces the production of superfluous products and conserves energy, maximizing energy efficiency.

Glossary

active site

enzyme's specific region to which the substrate binds

allosteric inhibition

inhibition by a binding event at a site different from the active site, which induces a conformational change and reduces the enzyme's affinity for its substrate

coenzyme

small organic molecule, such as a vitamin or its derivative, which is required to enhance an enzyme's activity

cofactor

inorganic ion, such as iron and magnesium ions, required for optimal enzyme activity regulation

competitive inhibition

type of inhibition in which the inhibitor competes with the substrate molecule by binding to the enzyme's active site

denature

process that changes a substance's natural properties

feedback inhibition

a product's effect of a reaction sequence to decrease its further production by inhibiting the first enzyme's activity in the pathway that produces it

induced fit

dynamic fit between the enzyme and its substrate, in which both components modify their structures to allow for ideal binding

substrate

molecule on which the enzyme acts

Energy in Living Systems

By the end of this section, you will be able to do the following:

- Discuss the importance of electrons in the transfer of energy in living systems
- Explain how ATP is used by cells as an energy source

Energy production within a cell involves many coordinated chemical pathways. Most of these pathways are combinations of oxidation and reduction reactions, which occur at the same time. An oxidation reaction strips an electron from an atom in a compound, and the addition of this electron to another compound is a reduction reaction. Because oxidation and reduction usually occur together, these pairs of reactions are called oxidation reduction reactions, or **redox reactions**.

Electrons and Energy

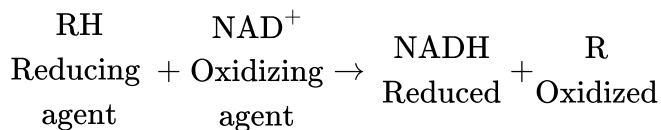
The removal of an electron from a molecule (oxidizing it), results in a decrease in potential energy in the oxidized compound. The electron (sometimes as part of a hydrogen atom) does not remain unbonded, however, in the cytoplasm of a cell. Rather, the electron is shifted to a second compound, reducing the second compound. *The shift of an electron from one compound to another removes some potential energy from the first compound (the oxidized compound) and increases the potential energy of the second compound (the reduced compound).* The transfer of electrons between molecules is important because most of the energy stored in atoms and used to fuel cell functions is in the form of high-energy electrons. The transfer of energy in the form of high-energy electrons allows the cell to transfer and use energy in an incremental fashion—in small packages rather than in a single, destructive burst. This chapter focuses on the extraction of energy from food; you will see that as you track the path of the transfers, you are tracking the path of electrons moving through metabolic pathways.

Electron Carriers

In living systems, a small class of compounds functions as electron shuttles: they bind and carry high-energy electrons between compounds in biochemical pathways. The principal electron carriers we will consider are derived from the B vitamin group and are derivatives of nucleotides. These compounds can be easily reduced (that is, they accept electrons) or oxidized (they lose electrons). Nicotinamide adenine dinucleotide (NAD) ([link](#)) is derived from vitamin B₃, niacin. NAD⁺ is the oxidized form of the molecule; NADH is the reduced form of the molecule after it has accepted two electrons and a proton (which together are the equivalent of a hydrogen atom with an extra electron). Note that if a compound has an “H” on it, it is generally reduced (e.g., NADH is the reduced form of NAD).

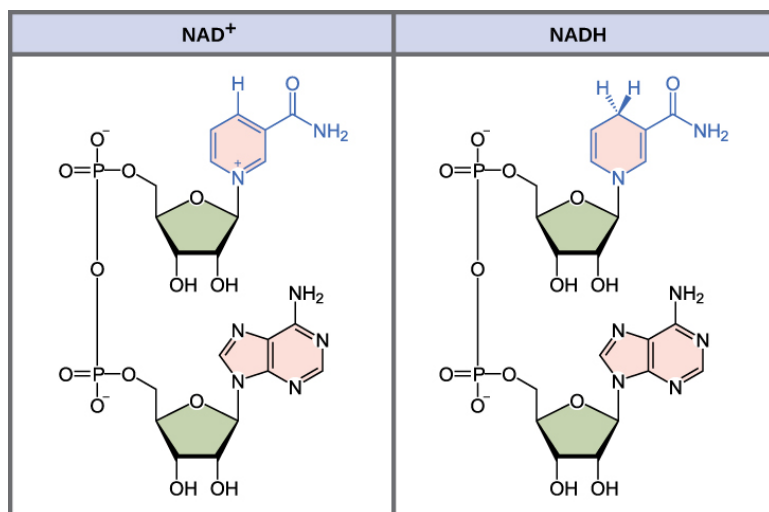
NAD⁺ can accept electrons from an organic molecule according to the general equation:

Equation:



When electrons are added to a compound, *it is reduced*. A compound that reduces another is called a reducing agent. In the above equation, RH is a reducing agent, and NAD^+ is reduced to NADH. When electrons are removed from a compound, *it is oxidized*. A compound that oxidizes another is called an oxidizing agent. In the above equation, NAD^+ is an oxidizing agent, and RH is oxidized to R.

Similarly, flavin adenine dinucleotide (FAD^+) is derived from vitamin B₂, also called riboflavin. Its reduced form is FADH_2 . A second variation of NAD, NADP, contains an extra phosphate group. Both NAD^+ and FAD^+ are extensively used in energy extraction from sugars, and NADP plays an important role in anabolic reactions and photosynthesis in plants.



The oxidized form of the electron carrier (NAD^+) is shown on the left, and the reduced form (NADH) is shown on the right. The nitrogenous base in NADH has one more hydrogen ion and two more electrons than in NAD^+ .

ATP in Living Systems

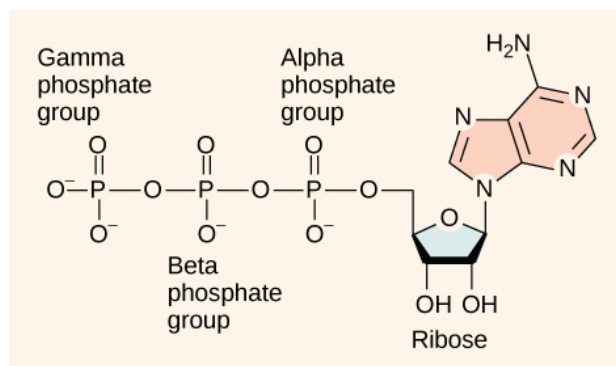
A living cell cannot store significant amounts of free energy. Excess free energy would result in an increase of heat in the cell, which would result in excessive thermal motion that could damage and then destroy the cell. Rather, a cell must be able to handle that energy in a way that enables the cell to store energy safely and release it for use only as needed. Living cells accomplish this by using the compound adenosine triphosphate (ATP). ATP is often called the “energy currency” of the cell, and, like currency, this versatile compound can be used to fill any energy need of the cell. How? It functions similarly to a rechargeable battery.

When ATP is broken down, usually by the removal of its terminal phosphate group, energy is released. The energy is used to do work by the cell, usually when the released phosphate binds to

another molecule, thereby activating it. For example, in the mechanical work of muscle contraction, ATP supplies the energy to move the contractile muscle proteins. Recall the active transport work of the sodium-potassium pump in cell membranes. ATP alters the structure of the integral protein that functions as the pump, changing its affinity for sodium and potassium. In this way, the cell performs work, pumping ions against their electrochemical gradients.

ATP Structure and Function

At the heart of ATP is a molecule of adenosine monophosphate (AMP), which is composed of an adenine molecule bonded to a ribose molecule and to a single phosphate group ([\[link\]](#)). Ribose is a five-carbon sugar found in RNA, and AMP is one of the nucleotides in RNA. The addition of a second phosphate group to this core molecule results in the formation of adenosine diphosphate (ADP); the addition of a third phosphate group forms adenosine triphosphate (ATP).



ATP (adenosine triphosphate) has three phosphate groups that can be removed by hydrolysis (addition of H₂O) to form ADP (adenosine diphosphate) or AMP (adenosine monophosphate). The negative charges on the phosphate group naturally repel each other, requiring energy to bond them together and releasing energy when these bonds are broken.

The addition of a phosphate group to a molecule requires energy. Phosphate groups are negatively charged and thus repel one another when they are arranged in series, as they are in ADP and ATP. This repulsion makes the ADP and ATP molecules inherently unstable. The release of one or two phosphate groups from ATP, a process called **dephosphorylation**, releases energy.

Energy from ATP

Hydrolysis is the process of breaking complex macromolecules apart. During hydrolysis, water is split, or lysed, and the resulting hydrogen atom (H^+) and a hydroxyl group (OH^-), or *hydroxide*, are added to the larger molecule. The hydrolysis of ATP produces ADP, together with an inorganic phosphate ion (P_i), and the release of free energy. To carry out life processes, ATP is continuously broken down into ADP, and like a rechargeable battery, ADP is continuously regenerated into ATP by the reattachment of a third phosphate group. Water, which was broken down into its hydrogen atom and hydroxyl group (hydroxide) during ATP hydrolysis, is regenerated when a third phosphate is added to the ADP molecule, reforming ATP.

Obviously, energy must be infused into the system to regenerate ATP. Where does this energy come from? In nearly every living thing on Earth, the energy comes from the metabolism of glucose, fructose, or galactose, all isomers with the chemical formula $C_6H_{12}O_6$ but different molecular configurations. In this way, ATP is a direct link between the limited set of exergonic pathways of glucose catabolism and the multitude of endergonic pathways that power living cells.

Phosphorylation

Recall that, in some chemical reactions, enzymes may bind to several substrates that react with each other on the enzyme, forming an intermediate complex. An intermediate complex is a temporary structure, and it allows one of the substrates (such as ATP) and reactants to more readily react with each other; in reactions involving ATP, ATP is one of the substrates and ADP is a product. During an endergonic chemical reaction, ATP forms an intermediate complex with the substrate and enzyme in the reaction. This intermediate complex allows the ATP to transfer its third phosphate group, with its energy, to the substrate, a process called phosphorylation.

Phosphorylation refers to the addition of the phosphate ($\sim P$). This is illustrated by the following generic reaction, in which A and B represent two different substrates:

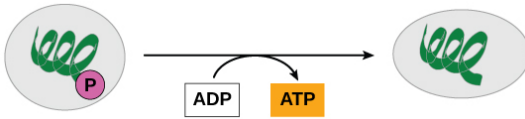
Equation:



When the intermediate complex breaks apart, the energy is used to modify the substrate and convert it into a product of the reaction. The ADP molecule and a free phosphate ion are released into the medium and are available for recycling through cell metabolism.

Substrate Phosphorylation

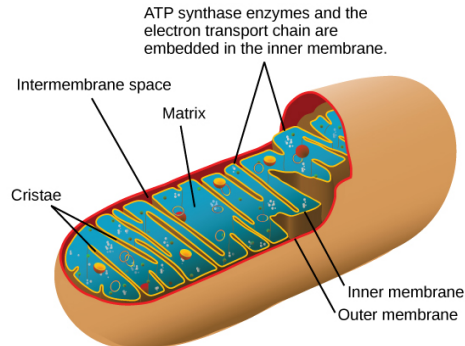
ATP is generated through two mechanisms during the breakdown of glucose. A few ATP molecules are generated (that is, regenerated from ADP) as a direct result of the chemical reactions that occur in the catabolic pathways. A phosphate group is removed from an intermediate reactant in the pathway, and the free energy of the reaction is used to add the third phosphate to an available ADP molecule, producing ATP ([link](#)). This very direct method of phosphorylation is called **substrate-level phosphorylation**.



In phosphorylation reactions, the gamma (third) phosphate of ATP is attached to a protein.

Oxidative Phosphorylation

Most of the ATP generated during glucose catabolism, however, is derived from a much more complex process, chemiosmosis, which takes place in mitochondria ([link](#)) within a eukaryotic cell or the plasma membrane of a prokaryotic cell. **Chemiosmosis**, a process of ATP production in cellular metabolism, is used to generate 90 percent of the ATP made during glucose catabolism and is also the method used in the light reactions of photosynthesis to harness the energy of sunlight. The production of ATP using the process of chemiosmosis is called **oxidative phosphorylation** because of the involvement of oxygen in the process.



In eukaryotes, oxidative phosphorylation takes place in mitochondria. In prokaryotes, this process takes place in the plasma membrane. (Credit: modification of work by Mariana Ruiz Villareal)

Note:

Career Connections

Mitochondrial Disease Physician

What happens when the critical reactions of cellular respiration do not proceed correctly? This may happen in mitochondrial diseases, which are genetic disorders of metabolism. Mitochondrial disorders can arise from mutations in nuclear or mitochondrial DNA, and they result in the production of less energy than is normal in body cells. In type 2 diabetes, for instance, the oxidation efficiency of NADH is reduced, impacting oxidative phosphorylation but not the other steps of respiration. Symptoms of mitochondrial diseases can include muscle weakness, lack of coordination, stroke-like episodes, and loss of vision and hearing. Most affected people are diagnosed in childhood, although there are some adult-onset diseases. Identifying and treating mitochondrial disorders is a specialized medical field. The educational preparation for this profession requires a college education, followed by medical school with a specialization in medical genetics. Medical geneticists can be board certified by the American Board of Medical Genetics and go on to become associated with professional organizations devoted to the study of mitochondrial diseases, such as the Mitochondrial Medicine Society and the Society for Inherited Metabolic Disorders.

Section Summary

ATP functions as the energy currency for cells. It allows the cell to store energy briefly and transport it within the cell to support endergonic chemical reactions. The structure of ATP is that of an RNA nucleotide with three phosphates attached. As ATP is used for energy, a phosphate group or two are detached, and either ADP or AMP is produced. Energy derived from glucose catabolism is used to convert ADP into ATP. When ATP is used in a reaction, the third phosphate is temporarily attached to a substrate in a process called phosphorylation. The two processes of ATP regeneration that are used in conjunction with glucose catabolism are substrate-level phosphorylation and oxidative phosphorylation through the process of chemiosmosis.

Review Questions

Exercise:

Problem: The energy currency used by cells is _____.

- a. ATP
- b. ADP
- c. AMP
- d. adenosine

Solution:

A

Exercise:

Problem: A reducing chemical reaction _____.

- a. reduces the compound to a simpler form
- b. adds an electron to the substrate
- c. removes a hydrogen atom from the substrate
- d. is a catabolic reaction

Solution:

B

Critical Thinking Questions

Exercise:

Problem:

Why is it beneficial for cells to use ATP rather than energy directly from the bonds of carbohydrates? What are the greatest drawbacks to harnessing energy directly from the bonds of several different compounds?

Solution:

ATP provides the cell with a way to handle energy in an efficient manner. The molecule can be charged, stored, and used as needed. Moreover, the energy from hydrolyzing ATP is delivered as a consistent amount. Harvesting energy from the bonds of several different compounds would result in energy deliveries of different quantities.

Glossary

chemiosmosis

process in which there is a production of adenosine triphosphate (ATP) in cellular metabolism by the involvement of a proton gradient across a membrane

dephosphorylation

removal of a phosphate group from a molecule

oxidative phosphorylation

production of ATP using the process of chemiosmosis in the presence of oxygen

phosphorylation

addition of a high-energy phosphate to a compound, usually a metabolic intermediate, a protein, or ADP

redox reaction

chemical reaction that consists of the coupling of an oxidation reaction and a reduction reaction

substrate-level phosphorylation

production of ATP from ADP using the excess energy from a chemical reaction and a phosphate group from a reactant

Glycolysis

By the end of this section, you will be able to do the following:

- Describe the overall result in terms of molecules produced during the chemical breakdown of glucose by glycolysis
- Compare the output of glycolysis in terms of ATP molecules and NADH molecules produced

As you have read, nearly all of the energy used by living cells comes to them in the bonds of the sugar glucose. **Glycolysis** is the first step in the breakdown of glucose to extract energy for cellular metabolism. In fact, nearly all living organisms carry out glycolysis as part of their metabolism. The process does not use oxygen directly and therefore is termed **anaerobic**. Glycolysis takes place in the cytoplasm of both prokaryotic and eukaryotic cells. Glucose enters heterotrophic cells in two ways. One method is through secondary active transport in which the transport takes place against the glucose concentration gradient. The other mechanism uses a group of integral proteins called **GLUT proteins**, also known as glucose transporter proteins. These transporters assist in the facilitated diffusion of glucose.

Glycolysis begins with the six-carbon ring-shaped structure of a single glucose molecule and ends with two molecules of a three-carbon sugar called **pyruvate**. Glycolysis consists of two distinct phases. The first part of the glycolysis pathway traps the glucose molecule in the cell and uses energy to modify it so that the six-carbon sugar molecule can be split evenly into the two three-carbon molecules. The second part of glycolysis extracts energy from the molecules and stores it in the form of ATP and NADH—remember: this is the reduced form of NAD.

First Half of Glycolysis (Energy-Requiring Steps)

Step 1. The first step in glycolysis ([\[link\]](#)) is catalyzed by hexokinase, an enzyme with broad specificity that catalyzes the phosphorylation of six-carbon sugars. Hexokinase phosphorylates glucose using ATP as the source of the phosphate, producing glucose-6-phosphate, a more reactive form of glucose. This reaction prevents the phosphorylated glucose molecule from

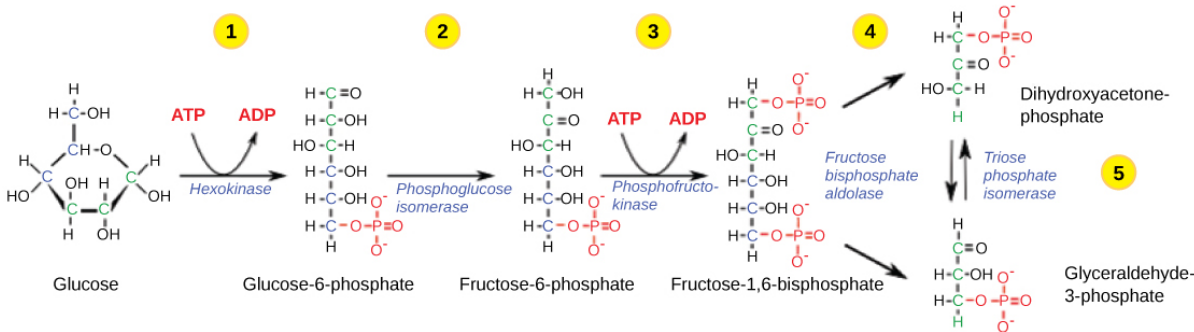
continuing to interact with the GLUT proteins, and it can no longer leave the cell because the negatively charged phosphate will not allow it to cross the hydrophobic interior of the plasma membrane.

Step 2. In the second step of glycolysis, an isomerase converts glucose-6-phosphate into one of its isomers, fructose-6-phosphate (this isomer has a phosphate attached at the location of the sixth carbon of the ring). An **isomerase** is an enzyme that catalyzes the conversion of a molecule into one of its isomers. (This change from phosphoglucose to phosphofructose allows the eventual split of the sugar into two three-carbon molecules.)

Step 3. The third step is the phosphorylation of fructose-6-phosphate, catalyzed by the enzyme phosphofructokinase. A second ATP molecule donates a high-energy phosphate to fructose-6-phosphate, producing fructose-1,6-bisphosphate. In this pathway, phosphofructokinase is a rate-limiting enzyme. It is active when the concentration of ADP is high; it is less active when ADP levels are low and the concentration of ATP is high. Thus, if there is “sufficient” ATP in the system, the pathway slows down. This is a type of end product inhibition, since ATP is the end product of glucose catabolism.

Step 4. The newly added high-energy phosphates further destabilize fructose-1,6-bisphosphate. The fourth step in glycolysis employs an enzyme, aldolase, to cleave fructose-1,6-bisphosphate into two three-carbon isomers: dihydroxyacetone phosphate and glyceraldehyde-3-phosphate.

Step 5. In the fifth step, an isomerase transforms the dihydroxyacetone-phosphate into its isomer, glyceraldehyde-3-phosphate. Thus, the pathway will continue with two molecules of a glyceraldehyde-3-phosphate. At this point in the pathway, there is a net investment of energy from two ATP molecules in the breakdown of one glucose molecule.

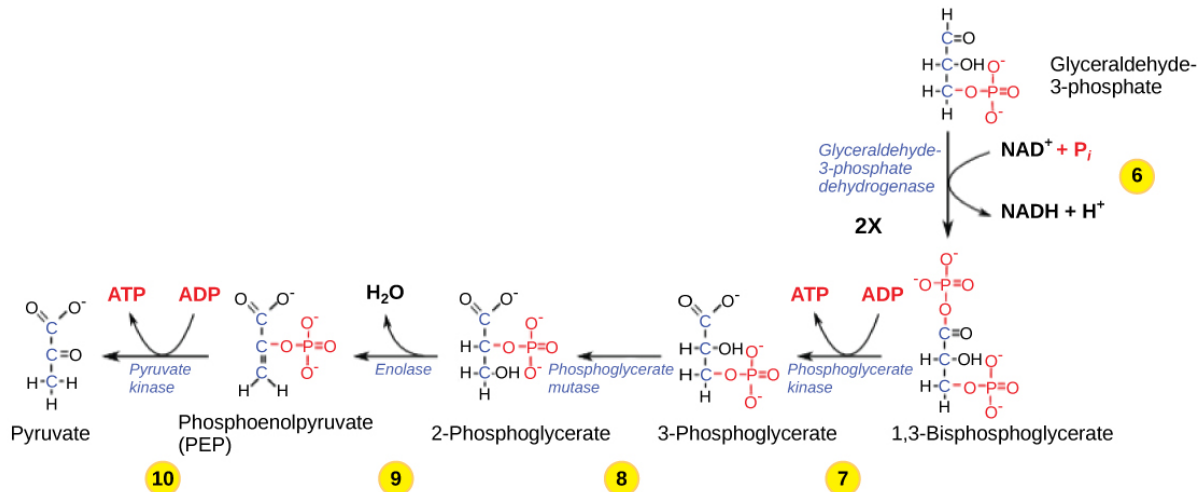


The first half of glycolysis uses two ATP molecules in the phosphorylation of glucose, which is then split into two three-carbon molecules.

Second Half of Glycolysis (Energy-Releasing Steps)

So far, glycolysis has cost the cell two ATP molecules and produced two small, three-carbon sugar molecules. Both of these molecules will proceed through the second half of the pathway, and sufficient energy will be extracted to pay back the two ATP molecules used as an initial investment and produce a profit for the cell of two additional ATP molecules and two even higher-energy NADH molecules.

Step 6. The sixth step in glycolysis ([\[link\]](#)) oxidizes the sugar (glyceraldehyde-3-phosphate), extracting high-energy electrons, which are picked up by the electron carrier NAD^+ , producing NADH. The sugar is then phosphorylated by the addition of a second phosphate group, producing 1,3-bisphosphoglycerate. Note that the second phosphate group does not require another ATP molecule.



The second half of glycolysis involves phosphorylation without ATP investment (step 6) and produces two NADH and four ATP molecules per glucose.

Here again is a potential limiting factor for this pathway. The continuation of the reaction depends upon the availability of the oxidized form of the electron carrier, NAD⁺. Thus, NADH must be continuously oxidized back into NAD⁺ in order to keep this step going. If NAD⁺ is not available, the second half of glycolysis slows down or stops. If oxygen is available in the system, the NADH will be oxidized readily, though indirectly, and the high-energy electrons from the hydrogen released in this process will be used to produce ATP. In an environment without oxygen, an alternate pathway (fermentation) can provide the oxidation of NADH to NAD⁺.

Step 7. In the seventh step, catalyzed by phosphoglycerate kinase (an enzyme named for the reverse reaction), 1,3-bisphosphoglycerate donates a high-energy phosphate to ADP, forming one molecule of ATP. (This is an example of substrate-level phosphorylation.) A carbonyl group on the 1,3-bisphosphoglycerate is oxidized to a carboxyl group, and 3-phosphoglycerate is formed.

Step 8. In the eighth step, the remaining phosphate group in 3-phosphoglycerate moves from the third carbon to the second carbon,

producing 2-phosphoglycerate (an isomer of 3-phosphoglycerate). The enzyme catalyzing this step is a mutase (isomerase).

Step 9. Enolase catalyzes the ninth step. This enzyme causes 2-phosphoglycerate to lose water from its structure; this is a dehydration reaction, resulting in the formation of a double bond that increases the potential energy in the remaining phosphate bond and produces phosphoenolpyruvate (PEP).

Step 10. The last step in glycolysis is catalyzed by the enzyme pyruvate kinase (the enzyme in this case is named for the reverse reaction of pyruvate's conversion into PEP) and results in the production of a second ATP molecule by substrate-level phosphorylation and the compound pyruvic acid (or its salt form, pyruvate). Many enzymes in enzymatic pathways are named for the reverse reactions, since the enzyme can catalyze both forward and reverse reactions (these may have been described initially by the reverse reaction that takes place in vitro, under nonphysiological conditions).

Note:

Link to Learning

Gain a better understanding of the breakdown of glucose by glycolysis by visiting this [site](#) to see the process in action.

Outcomes of Glycolysis

Glycolysis begins with glucose and produces two pyruvate molecules, four new ATP molecules, and two molecules of NADH. (Note: two ATP molecules are used in the first half of the pathway to prepare the six-carbon ring for cleavage, so the cell has a *net gain of two ATP molecules* and two NADH molecules for its use). If the cell cannot catabolize the pyruvate molecules further, it will harvest only two ATP molecules from one molecule of glucose. Mature mammalian red blood cells do not have mitochondria and thus are not capable of **aerobic respiration**—the process

in which organisms convert energy in the presence of oxygen—and glycolysis is their sole source of ATP. If glycolysis is interrupted, these cells lose their ability to maintain their sodium-potassium pumps, and eventually, they die.

The last step in glycolysis will not occur if pyruvate kinase, the enzyme that catalyzes the formation of pyruvate, is not available in sufficient quantities. In this situation, the entire glycolysis pathway will proceed, but only two ATP molecules will be made in the second half. Thus, pyruvate kinase is a rate-limiting enzyme for glycolysis.

Section Summary

Glycolysis is the first pathway within the cytoplasm used in the breakdown of glucose to extract energy. It was probably one of the earliest metabolic pathways to evolve and is used by nearly all of the organisms on Earth. Glycolysis consists of two parts: The first part prepares the six-carbon ring of glucose for cleavage into two three-carbon sugars. ATP is invested in the process during this half to energize the separation. The second half of glycolysis extracts ATP and high-energy electrons from hydrogen atoms and attaches them to NAD^+ . Two ATP molecules are invested in the first half and four ATP molecules are formed by substrate phosphorylation during the second half. This produces a net gain of two ATP and two NADH molecules for the cell.

Review Questions

Exercise:

Problem: During the second half of glycolysis, what occurs?

- a. ATP is used up.
 - b. Fructose is split in two.
 - c. ATP is made.
 - d. Glucose becomes fructose.
-

Solution:

C

Critical Thinking Questions

Exercise:**Problem:**

Nearly all organisms on Earth carry out some form of glycolysis. How does this fact support or not support the assertion that glycolysis is one of the oldest metabolic pathways?

Solution:

If glycolysis evolved relatively late, it likely would not be as universal in organisms as it is. It probably evolved in very primitive organisms and persisted, with the addition of other pathways of carbohydrate metabolism that evolved later.

Exercise:**Problem:**

Because they lose their mitochondria during development, red blood cells cannot perform aerobic respiration; however, they do perform glycolysis in the cytoplasm. Why do all cells need an energy source, and what would happen if glycolysis were blocked in a red blood cell?

Solution:

All cells must consume energy to carry out basic functions, such as pumping ions across membranes. A red blood cell would lose its membrane potential if glycolysis were blocked, and it would eventually die.

Glossary

aerobic respiration

process in which organisms convert energy in the presence of oxygen

anaerobic

process that does not use oxygen

glycolysis

process of breaking glucose into two three-carbon molecules with the production of ATP and NADH

isomerase

enzyme that converts a molecule into its isomer

pyruvate

three-carbon sugar that can be decarboxylated and oxidized to make acetyl CoA, which enters the citric acid cycle under aerobic conditions; the end product of glycolysis

Oxidative Phosphorylation

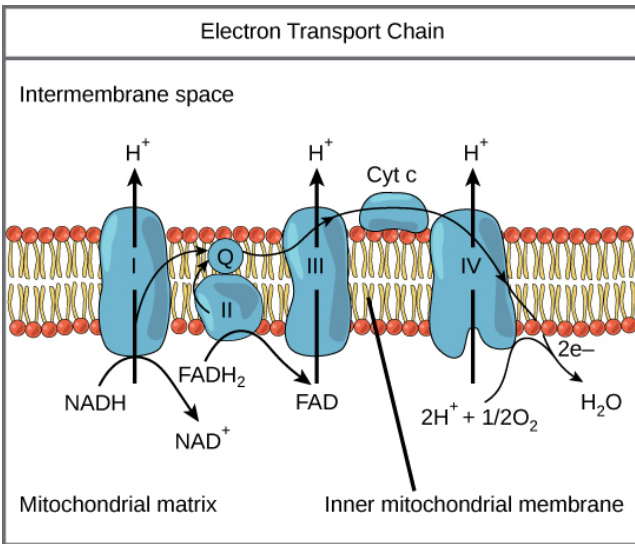
By the end of this section, you will be able to do the following:

- Describe how electrons move through the electron transport chain and explain what happens to their energy levels during this process
- Explain how a proton (H^+) gradient is established and maintained by the electron transport chain

You have just read about two pathways in glucose catabolism—glycolysis and the citric acid cycle—that generate ATP. Most of the ATP generated during the aerobic catabolism of glucose, however, is not generated directly from these pathways. Instead, it is derived from a process that begins by moving electrons through a series of electron carriers that undergo redox reactions. This process causes hydrogen ions to accumulate within the intermembranous space. Therefore, a concentration gradient forms in which hydrogen ions diffuse out of the intermembranous space into the mitochondrial matrix by passing through ATP synthase. The current of hydrogen ions powers the catalytic action of ATP synthase, which phosphorylates ADP, producing ATP.

Electron Transport Chain

The electron transport chain ([\[link\]](#)) is the last component of aerobic respiration and is the only part of glucose metabolism that uses atmospheric oxygen. Oxygen continuously diffuses into plant tissues (typically through stomata), as well as into fungi and bacteria; however, in animals, oxygen enters the body through a variety of respiratory systems. Electron transport is a series of redox reactions that resembles a relay race or bucket brigade in that electrons are passed rapidly from one component to the next, to the endpoint of the chain where the electrons reduce molecular oxygen and, along with associated protons, produces water. There are four complexes composed of proteins, labeled I through IV in [\[link\]](#), and the aggregation of these four complexes, together with associated mobile, accessory electron carriers, is called the **electron transport chain**. The electron transport chain is present with multiple copies in the inner mitochondrial membrane of eukaryotes and within the plasma membrane of prokaryotes.



The electron transport chain is a series of electron transporters embedded in the inner mitochondrial membrane that shuttles electrons from NADH and FADH₂ to molecular oxygen. In the process, protons are pumped from the mitochondrial matrix to the intermembrane space, and oxygen is reduced to form water.

Complex I

First, two electrons are carried to the first complex via NADH. This complex, labeled **I**, is composed of flavin mononucleotide (FMN) and an iron-sulfur (Fe-S)-containing protein. FMN, which is derived from vitamin B₂ (also called riboflavin), is one of several prosthetic groups or cofactors in the electron transport chain. A **prosthetic group** is a nonprotein molecule required for the activity of a protein. Prosthetic groups are organic or inorganic, nonpeptide molecules bound to a protein that facilitate its function. Prosthetic groups include coenzymes, which are the prosthetic

groups of enzymes. The enzyme in complex I is NADH dehydrogenase and is a very large protein, containing 45 amino acid chains. Complex I can pump four hydrogen ions across the membrane from the matrix into the intermembrane space, and it is in this way that the hydrogen ion gradient is established and maintained between the two compartments separated by the inner mitochondrial membrane.

Q and Complex II

Complex II directly receives FADH_2 —which does not pass through complex I. The compound connecting the first and second complexes to the third is **ubiquinone** B. The Q molecule is lipid soluble and freely moves through the hydrophobic core of the membrane. Once it is reduced (QH_2), ubiquinone delivers its electrons to the next complex in the electron transport chain. Q receives the electrons derived from NADH from complex I, and the electrons derived from FADH_2 from complex II. This enzyme and FADH_2 form a small complex that delivers electrons directly to the electron transport chain, bypassing the first complex. Since these electrons bypass and thus do not energize the proton pump in the first complex, fewer ATP molecules are made from the FADH_2 electrons. *The number of ATP molecules ultimately obtained is directly proportional to the number of protons pumped across the inner mitochondrial membrane.*

Complex III

The third complex is composed of cytochrome b—another Fe-S protein, a Rieske center (2Fe-2S center), and cytochrome c proteins. This complex is also called cytochrome oxidoreductase. Cytochrome proteins have a prosthetic group of heme. The heme molecule is similar to the heme in hemoglobin, but it carries electrons, not oxygen. As a result, the iron ion at its core is reduced and oxidized as it passes the electrons, fluctuating between different oxidation states: Fe^{++} (reduced) and Fe^{+++} (oxidized). The heme molecules in the cytochromes have slightly different characteristics due to the effects of the different proteins binding to them,

giving slightly different characteristics to each complex. Complex III pumps protons through the membrane and passes its electrons to cytochrome c for transport to the fourth complex of proteins and enzymes. (Cytochrome c receives electrons from Q; however, whereas Q carries pairs of electrons, cytochrome c can accept only one at a time.)

Complex IV

The fourth complex is composed of cytochrome proteins c, a, and a_3 . This complex contains two heme groups (one in each of the two cytochromes, a, and a_3) and three copper ions (a pair of Cu_A and one Cu_B in cytochrome a_3). The cytochromes hold an oxygen molecule very tightly between the iron and copper ions until the oxygen is completely reduced by the gain of two electrons. The reduced oxygen then picks up two hydrogen ions from the surrounding medium to make water (H_2O). The removal of the hydrogen ions from the system contributes to the ion gradient that forms the foundation for the process of chemiosmosis.

Chemiosmosis

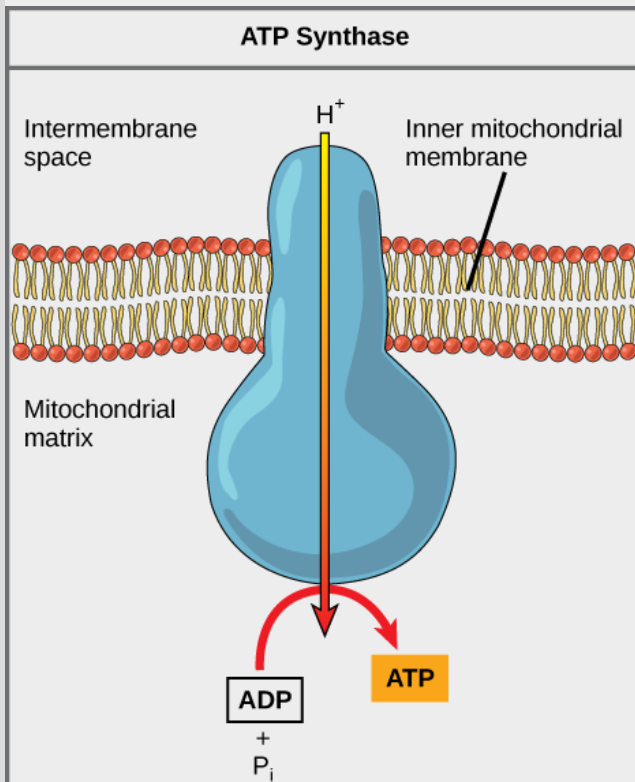
In chemiosmosis, the free energy from the series of redox reactions just described is used to pump hydrogen ions (protons) across the mitochondrial membrane. The uneven distribution of H^+ ions across the membrane establishes both concentration and electrical gradients (thus, an electrochemical gradient), owing to the hydrogen ions' positive charge and their aggregation on one side of the membrane.

If the membrane were continuously open to simple diffusion by the hydrogen ions, the ions would tend to diffuse back across into the matrix, driven by the concentrations producing their electrochemical gradient. Recall that many ions cannot diffuse through the nonpolar regions of phospholipid membranes without the aid of ion channels. Similarly, hydrogen ions in the matrix space can only pass through the inner mitochondrial membrane by an integral membrane protein called ATP synthase ([\[link\]](#)). This complex protein acts as a tiny generator, turned by

the force of the hydrogen ions diffusing through it, down their electrochemical gradient. The turning of parts of this molecular machine facilitates the addition of a phosphate to ADP, forming ATP, *using the potential energy of the hydrogen ion gradient.*

Note:

Visual Connection

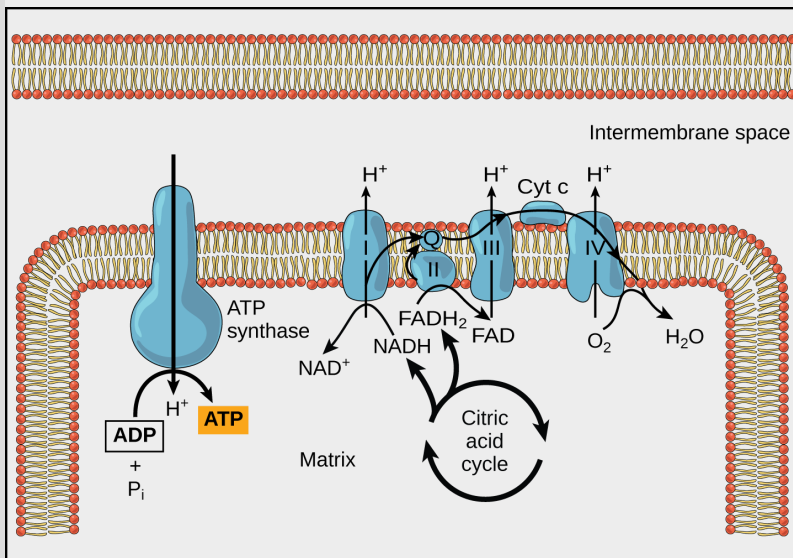


ATP synthase is a complex, molecular machine that uses a proton (H⁺) gradient to form ATP from ADP and inorganic phosphate (P_i). (Credit: modification of work by Klaus Hoffmeier)

Dinitrophenol (DNP) is an “uncoupler” that makes the inner mitochondrial membrane “leaky” to protons. It was used until 1938 as a weight-loss drug. What effect would you expect DNP to have on the change in pH across the inner mitochondrial membrane? Why do you think this might be an effective weight-loss drug?

Chemiosmosis ([\[link\]](#)) is used to generate 90 percent of the ATP made during aerobic glucose catabolism; it is also the method used in the light reactions of photosynthesis to harness the energy of sunlight in the process of photophosphorylation. Recall that the production of ATP using the process of chemiosmosis in mitochondria is called oxidative phosphorylation. The overall result of these reactions is the production of ATP from the energy of the electrons removed from hydrogen atoms. These atoms were originally part of a glucose molecule. At the end of the pathway, the electrons are used to reduce an oxygen molecule to oxygen ions. The extra electrons on the oxygen attract hydrogen ions (protons) from the surrounding medium, and water is formed. Thus, oxygen is the final electron acceptor in the electron transport chain.

Note:
Visual Connection



In oxidative phosphorylation, the pH gradient formed by the electron transport chain is used by ATP synthase to form ATP.

Cyanide inhibits cytochrome c oxidase, a component of the electron transport chain. If cyanide poisoning occurs, would you expect the pH of the intermembrane space to increase or decrease? What effect would cyanide have on ATP synthesis?

ATP Yield

The number of ATP molecules generated from the catabolism of glucose varies. For example, the number of hydrogen ions that the electron transport chain complexes can pump through the membrane varies between species. Another source of variance stems from the shuttle of electrons across the membranes of the mitochondria. (The NADH generated from glycolysis cannot easily enter mitochondria.) Thus, electrons are picked up on the inside of mitochondria by either NAD^+ or FAD^+ . As you have learned earlier, these FAD^+ molecules can transport fewer ions; consequently, fewer ATP molecules are generated when FAD^+ acts as a carrier. NAD^+ is used as the electron transporter in the liver and FAD^+ acts in the brain.

Another factor that affects the yield of ATP molecules generated from glucose is the fact that intermediate compounds in these pathways are also used for other purposes. Glucose catabolism connects with the pathways that build or break down all other biochemical compounds in cells, and the result is somewhat messier than the ideal situations described thus far. For example, sugars other than glucose are fed into the glycolytic pathway for energy extraction. In addition, the five-carbon sugars that form nucleic acids are made from intermediates in glycolysis. Certain nonessential amino acids can be made from intermediates of both glycolysis and the citric acid cycle. Lipids, such as cholesterol and triglycerides, are also made from intermediates in these pathways, and both amino acids and triglycerides are broken down for energy through these pathways. Overall, in living systems,

these pathways of glucose catabolism extract about 34 percent of the energy contained in glucose, with the remainder being released as heat.

Section Summary

The electron transport chain is the portion of aerobic respiration that uses free oxygen as the final electron acceptor of the electrons removed from the intermediate compounds in glucose catabolism. The electron transport chain is composed of four large, multiprotein complexes embedded in the inner mitochondrial membrane and two small diffusible electron carriers shuttling electrons between them. The electrons are passed through a series of redox reactions, with a small amount of free energy used at three points to transport hydrogen ions across a membrane. This process contributes to the gradient used in chemiosmosis. The electrons passing through the electron transport chain gradually lose energy. High-energy electrons donated to the chain by either NADH or FADH₂ complete the chain, as low-energy electrons reduce oxygen molecules and form water. The level of free energy of the electrons drops from about 60 kcal/mol in NADH or 45 kcal/mol in FADH₂ to about 0 kcal/mol in water. The end products of the electron transport chain are water and ATP. A number of intermediate compounds of the citric acid cycle can be diverted into the anabolism of other biochemical molecules, such as nonessential amino acids, sugars, and lipids. These same molecules can serve as energy sources for the glucose pathways.

Visual Connection Questions

Exercise:

Problem:

[\[link\]](#) Dinitrophenol (DNP) is an "uncoupler" that makes the inner mitochondrial membrane "leaky" to protons. It was used until 1938 as a weight-loss drug. What effect would you expect DNP to have on the change in pH across the inner mitochondrial membrane? Why do you think this might be an effective weight-loss drug?

Solution:

[\[link\]](#) After DNP poisoning, the electron transport chain can no longer form a proton gradient, and ATP synthase can no longer make ATP. DNP is an effective diet drug because it uncouples ATP synthesis; in other words, after taking it, a person obtains less energy out of the food he or she eats. Interestingly, one of the worst side effects of this drug is hyperthermia, or overheating of the body. Since ATP cannot be formed, the energy from electron transport is lost as heat.

Exercise:

Problem:

[\[link\]](#) Cyanide inhibits cytochrome c oxidase, a component of the electron transport chain. If cyanide poisoning occurs, would you expect the pH of the intermembrane space to increase or decrease? What effect would cyanide have on ATP synthesis?

Solution:

[\[link\]](#) After cyanide poisoning, the electron transport chain can no longer pump electrons into the intermembrane space. The pH of the intermembrane space would increase, the pH gradient would decrease, and ATP synthesis would stop.

Review Questions

Exercise:

Problem: What compound receives electrons from NADH?

- a. FMN
- b. ubiquinone
- c. cytochrome c_1
- d. oxygen

Solution:

A

Exercise:

Problem: Chemiosmosis involves _____.

- a. the movement of electrons across the cell membrane
- b. the movement of hydrogen atoms across a mitochondrial membrane
- c. the movement of hydrogen ions across a mitochondrial membrane
- d. the movement of glucose through the cell membrane

Solution:

C

Critical Thinking Questions

Exercise:

Problem:

How do the roles of ubiquinone and cytochrome c differ from the roles of the other components of the electron transport chain?

Solution:

Q and cytochrome c are transport molecules. Their function does not result directly in ATP synthesis in that they are not pumps. Moreover, Q is the only component of the electron transport chain that is not a protein. Ubiquinone and cytochrome c are small, mobile electron carriers, whereas the other components of the electron transport chain are large complexes anchored in the inner mitochondrial membrane.

Exercise:

Problem:

What accounts for the different number of ATP molecules that are formed through cellular respiration?

Solution:

Few tissues except muscle produce the maximum possible amount of ATP from nutrients. The intermediates are used to produce needed amino acids, fatty acids, cholesterol, and sugars for nucleic acids. When NADH is transported from the cytoplasm to the mitochondria, an active transport mechanism is used, which decreases the amount of ATP that can be made. The electron transport chain differs in composition between species, so different organisms will make different amounts of ATP using their electron transport chains.

Glossary**ATP synthase**

(also F1F0 ATP synthase) membrane-embedded protein complex that adds a phosphate to ADP with energy from protons diffusing through it

prosthetic group

(also prosthetic cofactor) molecule bound to a protein that facilitates the function of the protein

ubiquinone

soluble electron transporter in the electron transport chain that connects the first or second complex to the third

Metabolism without Oxygen

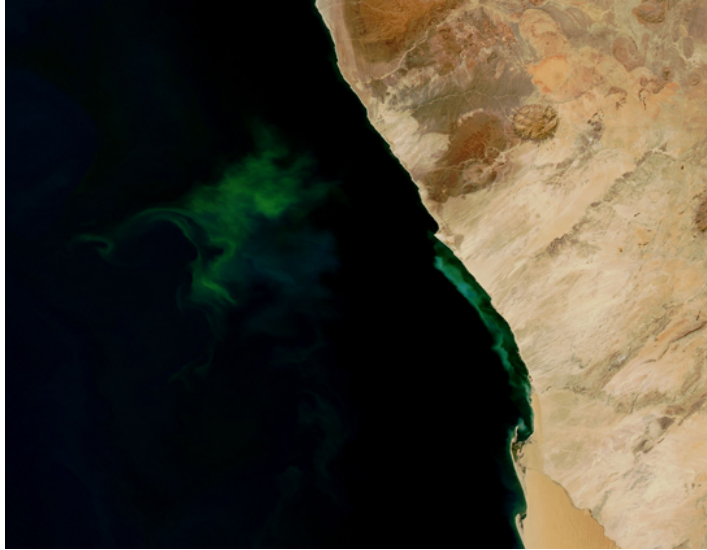
By the end of this section, you will be able to do the following:

- Discuss the fundamental difference between anaerobic cellular respiration and fermentation
- Describe the type of fermentation that readily occurs in animal cells and the conditions that initiate that fermentation

In aerobic respiration, the final electron acceptor is an oxygen molecule, O_2 . If aerobic respiration occurs, then ATP will be produced using the energy of high-energy electrons carried by NADH or $FADH_2$ to the electron transport chain. If aerobic respiration does not occur, NADH must be reoxidized to NAD^+ for reuse as an electron carrier for the glycolytic pathway to continue. How is this done? Some living systems use an organic molecule as the final electron acceptor. Processes that use an organic molecule to regenerate NAD^+ from NADH are collectively referred to as **fermentation**. In contrast, some living systems use an inorganic molecule as a final electron acceptor. Both methods are called **anaerobic cellular respiration**, in which organisms convert energy for their use in the absence of oxygen.

Anaerobic Cellular Respiration

Certain prokaryotes, including some species in the domains Bacteria and Archaea, use anaerobic respiration. For example, a group of archaeans called methanogens reduces carbon dioxide to methane to oxidize NADH. These microorganisms are found in soil and in the digestive tracts of ruminants, such as cows and sheep. Similarly, sulfate-reducing bacteria, most of which are anaerobic ([link](#)), reduce sulfate to hydrogen sulfide to regenerate NAD^+ from NADH.



The green color seen in these coastal waters is from an eruption of hydrogen sulfide–producing bacteria. These anaerobic, sulfate-reducing bacteria release hydrogen sulfide gas as they decompose algae in the water. (credit: modification of work by NASA/Jeff Schmaltz, MODIS Land Rapid Response Team at NASA GSFC, Visible Earth Catalog of NASA images)

Note:

Link to Learning

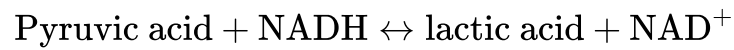
Visit this [site](#) to see anaerobic cellular respiration in action.

Lactic Acid Fermentation

The fermentation method used by animals and certain bacteria, such as those in yogurt, is lactic acid fermentation ([link](#)). This type of fermentation is used routinely in mammalian red blood cells, which do not have mitochondria, and in skeletal muscle that has an insufficient oxygen supply to allow aerobic respiration to continue (that is, in muscles used to the point of fatigue). In muscles, lactic acid accumulation must be removed by the blood circulation, and when the lactic acid loses a hydrogen, the

resulting lactate is brought to the liver for further metabolism. The chemical reactions of lactic acid fermentation are the following:

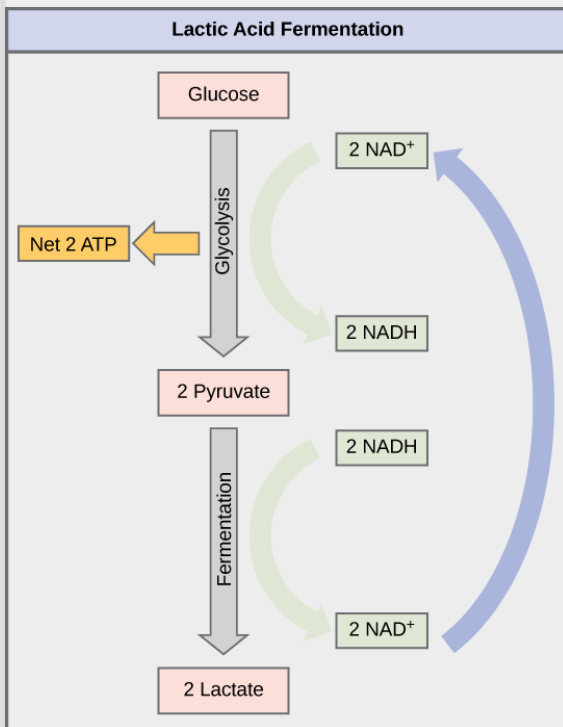
Equation:



The enzyme used in this reaction is lactate dehydrogenase (LDH). The reaction can proceed in either direction, but the reaction from left to right is inhibited by acidic conditions. Such lactic acid accumulation was once believed to cause muscle stiffness, fatigue, and soreness, although more recent research disputes this hypothesis. Once the lactic acid has been removed from the muscle and circulated to the liver, it can be reconverted into pyruvic acid and further catabolized for energy.

Note:

Visual Connection



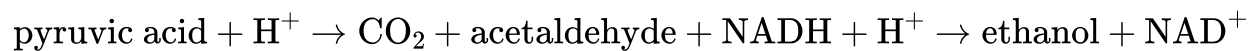
Lactic acid fermentation is common in muscle cells that have run out of oxygen.

Tremetol, a metabolic poison found in the white snakeroot plant, prevents the metabolism of lactate. When cows eat this plant, tremetol is concentrated in the milk they produce. Humans who consume the milk can become seriously ill. Symptoms of this disease, which include vomiting, abdominal pain, and tremors, become worse after exercise. Why do you think this is the case?

Alcohol Fermentation

Another familiar fermentation process is alcohol fermentation ([\[link\]](#)), which produces ethanol. The first chemical reaction of alcohol fermentation is the following (CO_2 does not participate in the second reaction):

Equation:



The first reaction is catalyzed by pyruvate decarboxylase, a cytoplasmic enzyme, with a coenzyme of thiamine pyrophosphate (TPP, derived from vitamin B_1 and also called thiamine). A carboxyl group is removed from pyruvic acid, releasing carbon dioxide as a gas. The loss of carbon dioxide reduces the size of the molecule by one carbon, producing acetaldehyde. The second reaction is catalyzed by alcohol dehydrogenase to oxidize NADH to NAD^+ and reduce acetaldehyde to ethanol. The fermentation of pyruvic acid by yeast produces the ethanol found in alcoholic beverages. Ethanol tolerance of yeast is variable, ranging from about 5 percent to 21 percent, depending on the yeast strain and environmental conditions.



Fermentation of grape juice into wine produces CO₂ as a byproduct. Fermentation tanks have valves so that the pressure inside the tanks created by the carbon dioxide produced can be released.

Other Types of Fermentation

Other fermentation methods take place in bacteria. We should note that many prokaryotes are *facultatively* anaerobic. This means that they can switch between aerobic respiration and fermentation, depending on the availability of free oxygen. Certain prokaryotes, such as *Clostridia*, are obligate anaerobes. Obligate anaerobes live and grow in the absence of molecular oxygen. Oxygen is a poison to these microorganisms and kills them on exposure. We should also note that all forms of fermentation, except lactic acid fermentation, produce gas. The production of particular types of gas is used as an indicator of the fermentation of specific carbohydrates, which plays a role in the laboratory identification of the bacteria. Various methods of fermentation are used by assorted organisms to ensure an adequate supply of NAD⁺ for the sixth step in glycolysis. Without these pathways, this step would not occur, and ATP could not be harvested from the breakdown of glucose.

Section Summary

If NADH cannot be oxidized through aerobic respiration, another electron acceptor is used. Most organisms will use some form of fermentation to accomplish the regeneration of NAD⁺, ensuring the continuation of glycolysis. The regeneration of NAD⁺ in fermentation is not accompanied by ATP production; therefore, the potential of NADH to produce ATP using an electron transport chain is not utilized.

Visual Connection Questions

Exercise:

Problem:

[\[link\]](#) Tremetol, a metabolic poison found in the white snake root plant, prevents the metabolism of lactate. When cows eat this plant, tremetol is concentrated in the milk they produce. Humans who consume the milk can become seriously ill. Symptoms of this disease, which include vomiting, abdominal pain, and tremors, become worse after exercise. Why do you think this is the case?

Solution:

[\[link\]](#) The illness is caused by lactate accumulation. Lactate levels rise after exercise, making the symptoms worse. Milk sickness is rare today but was common in the midwestern United States in the early 1800s.

Review Questions**Exercise:****Problem:**

Which of the following fermentation methods can occur in animal skeletal muscles?

- a. lactic acid fermentation
- b. alcohol fermentation
- c. mixed acid fermentation
- d. propionic fermentation

Solution:

A

Critical Thinking Questions**Exercise:****Problem:**

What is the primary difference between fermentation and anaerobic respiration?

Solution:

Fermentation uses glycolysis only. Anaerobic respiration uses all three parts of cellular respiration, including the parts in the mitochondria like the citric acid cycle and electron transport; it also uses a different final electron acceptor instead of oxygen gas.

Glossary

anaerobic cellular respiration

process in which organisms convert energy for their use in the absence of oxygen

fermentation

process of regenerating NAD^+ with either an inorganic or organic compound serving as the final electron acceptor; occurs in the absence of oxygen

Connections of Carbohydrate, Protein, and Lipid Metabolic Pathways

By the end of this section, you will be able to do the following:

- Discuss the ways in which carbohydrate metabolic pathways, glycolysis, and the citric acid cycle interrelate with protein and lipid metabolic pathways
- Explain why metabolic pathways are not considered closed systems

You have learned about the catabolism of glucose, which provides energy to living cells. But living things consume organic compounds other than glucose for food. How does a turkey sandwich end up as ATP in your cells? This happens because all of the catabolic pathways for carbohydrates, proteins, and lipids eventually connect into glycolysis and the citric acid cycle pathways (see [\[link\]](#)). Metabolic pathways should be thought of as porous and interconnecting—that is, substances enter from other pathways, and intermediates leave for other pathways. These pathways are not closed systems! Many of the substrates, intermediates, and products in a particular pathway are reactants in other pathways.

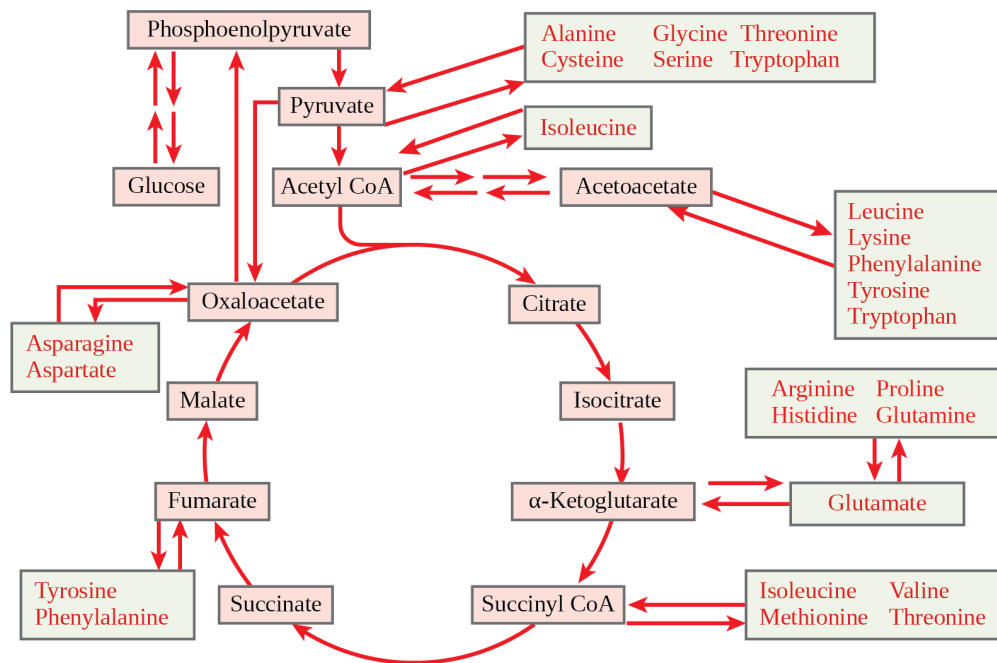
Connections of Other Sugars to Glucose Metabolism

Glycogen, a polymer of glucose, is an energy storage molecule in animals. When there is adequate ATP present, excess glucose is stored as glycogen in both liver and muscle cells. The glycogen will be hydrolyzed into glucose 1-phosphate monomers (G-1-P) if blood sugar levels drop. The presence of glycogen as a source of glucose allows ATP to be produced for a longer period of time during exercise. Glycogen is broken down into glucose-1-phosphate (G-1-P) and converted into glucose-6-phosphate (G-6-P) in both muscle and liver cells, and this product enters the glycolytic pathway.

Sucrose is a disaccharide with a molecule of glucose and a molecule of fructose bonded together with a glycosidic linkage. Fructose is one of the three “dietary” monosaccharides, along with glucose and galactose (part of the milk sugar disaccharide lactose), which are absorbed directly into the bloodstream during digestion. The catabolism of both fructose and galactose produces the same number of ATP molecules as glucose.

Connections of Proteins to Glucose Metabolism

Proteins are hydrolyzed by a variety of enzymes in cells. Most of the time, the amino acids are recycled into the synthesis of new proteins. If there are excess amino acids, however, or if the body is in a state of starvation, some amino acids will be shunted into the pathways of glucose catabolism ([link](#)). It is very important to note that each amino acid must have its amino group removed prior to entry into these pathways. The amino group is converted into ammonia. In mammals, the liver synthesizes urea from two ammonia molecules and a carbon dioxide molecule. Thus, urea is the principal waste product in mammals, produced from the nitrogen originating in amino acids, and it leaves the body in urine. It should be noted that amino acids can be synthesized from the intermediates and reactants in the cellular respiration cycle.

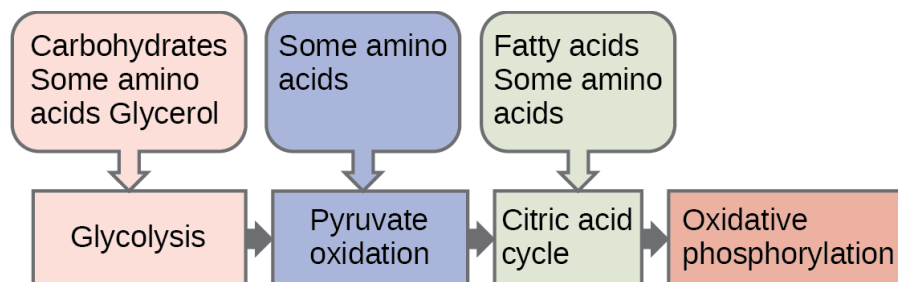


The carbon skeletons of certain amino acids (indicated in boxes) derived from proteins can feed into the citric acid cycle. (credit: modification of work by Mikael Häggström)

Connections of Lipid and Glucose Metabolisms

The lipids connected to the glucose pathway include cholesterol and triglycerides. Cholesterol is a lipid that contributes to cell membrane flexibility and is a precursor of steroid hormones. The synthesis of cholesterol starts with acetyl groups and proceeds in only one direction. The process cannot be reversed.

Triglycerides—made from the bonding of glycerol and three fatty acids—are a form of long-term energy storage in animals. Animals can make most of the fatty acids they need. Triglycerides can be both made and broken down through parts of the glucose catabolism pathways. Glycerol can be phosphorylated to glycerol-3-phosphate, which continues through glycolysis. Fatty acids are catabolized in a process called beta-oxidation, which takes place in the matrix of the mitochondria and converts their fatty acid chains into two-carbon units of acetyl groups. The acetyl groups are picked up by CoA to form acetyl CoA that proceeds into the citric acid cycle.



Glycogen from the liver and muscles, as well as other carbohydrates, hydrolyzed into glucose-1-phosphate, together with fats and proteins, can feed into the catabolic pathways for carbohydrates.

Note:**Evolution Connection****Pathways of Photosynthesis and Cellular Metabolism**

The processes of photosynthesis and cellular metabolism consist of several very complex pathways. It is generally thought that the first cells arose in an aqueous environment—a “soup” of nutrients—possibly on the surface of some porous clays, perhaps in warm marine environments. If these cells reproduced successfully and their numbers climbed steadily, it follows that the cells would begin to deplete the nutrients from the medium in which they lived as they shifted the nutrients into the components of their own bodies. This hypothetical situation would have resulted in natural selection favoring those organisms that could exist by using the nutrients that remained in their environment and by manipulating these nutrients into materials upon which they could survive. Selection would favor those organisms that could extract maximal value from the nutrients to which they had access.

An early form of photosynthesis developed that harnessed the sun’s energy using water as a source of hydrogen atoms, but this pathway did not produce free oxygen (anoxygenic photosynthesis). (Another type of anoxygenic photosynthesis did not produce free oxygen because it did not use water as the source of hydrogen ions; instead, it used materials such as hydrogen sulfide and consequently produced sulfur). It is thought that glycolysis developed at this time and could take advantage of the simple sugars being produced but that these reactions were unable to fully extract the energy stored in the carbohydrates. The development of glycolysis probably predated the evolution of photosynthesis, as it was well suited to extract energy from materials spontaneously accumulating in the “primeval soup.” A later form of photosynthesis used water as a source of electrons and hydrogen and generated free oxygen. Over time, the atmosphere became oxygenated, but not before the oxygen released oxidized metals in the ocean and created a “rust” layer in the sediment, permitting the dating of the rise of the first oxygenic photosynthesizers. Living things adapted to exploit this new atmosphere that allowed aerobic respiration as we know it to evolve. When the full process of oxygenic photosynthesis developed and the atmosphere became oxygenated, cells were finally able to use the oxygen expelled by photosynthesis to extract considerably more energy

from the sugar molecules using the citric acid cycle and oxidative phosphorylation.

Section Summary

The breakdown and synthesis of carbohydrates, proteins, and lipids connect with the pathways of glucose catabolism. The simple sugars are galactose, fructose, glycogen, and pentose. These are catabolized during glycolysis. The amino acids from proteins connect with glucose catabolism through pyruvate, acetyl CoA, and components of the citric acid cycle. Cholesterol synthesis starts with acetyl groups, and the components of triglycerides come from glycerol-3-phosphate from glycolysis and acetyl groups produced in the mitochondria from pyruvate.

Review Questions

Exercise:

Problem: A major connection for sugars in glycolysis is _____.

- a. glucose-6-phosphate
- b. fructose-1,6-bisphosphate
- c. dihydroxyacetone phosphate
- d. phosphoenolpyruvate

Solution:

A

Exercise:

Problem: Beta-oxidation is _____.

- a. the breakdown of sugars

- b. the assembly of sugars
- c. the breakdown of fatty acids
- d. the removal of amino groups from amino acids

Solution:

C

Critical Thinking Questions

Exercise:

Problem:

Would you describe metabolic pathways as inherently wasteful or inherently economical? Why?

Solution:

They are very economical. The substrates, intermediates, and products move between pathways and do so in response to finely tuned feedback inhibition loops that keep metabolism balanced overall. Intermediates in one pathway may occur in another, and they can move from one pathway to another fluidly in response to the needs of the cell.

Regulation of Cellular Respiration

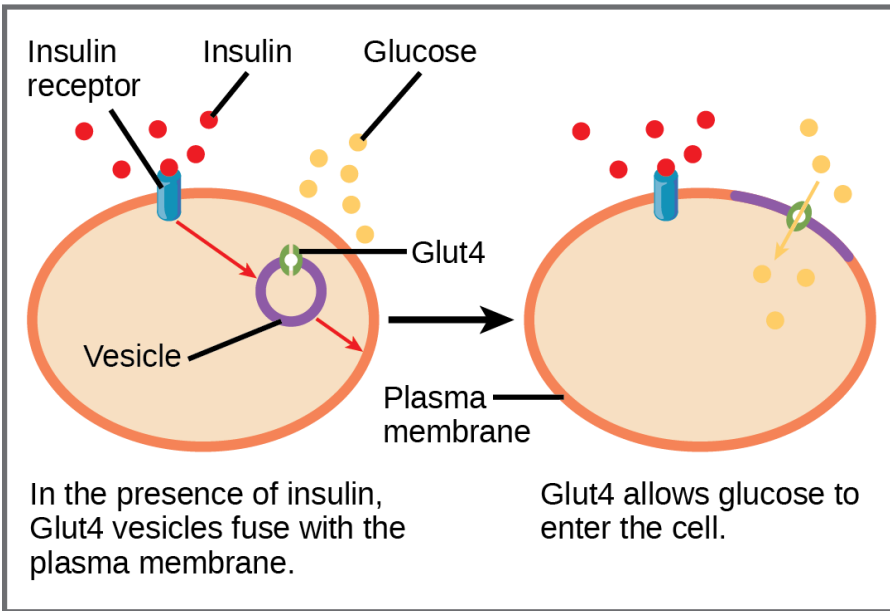
By the end of this section, you will be able to do the following:

- Describe how feedback inhibition would affect the production of an intermediate or product in a pathway
- Identify the mechanism that controls the rate of the transport of electrons through the electron transport chain

Cellular respiration must be regulated in order to provide balanced amounts of energy in the form of ATP. The cell also must generate a number of intermediate compounds that are used in the anabolism and catabolism of macromolecules. Without controls, metabolic reactions would quickly come to a standstill as the forward and backward reactions reached a state of equilibrium. Resources would be used inappropriately. A cell does not need the maximum amount of ATP that it can make all the time: At times, the cell needs to shunt some of the intermediates to pathways for amino acid, protein, glycogen, lipid, and nucleic acid production. In short, the cell needs to control its metabolism.

Regulatory Mechanisms

A variety of mechanisms is used to control cellular respiration. Some type of control exists at each stage of glucose metabolism. Access of glucose to the cell can be regulated using the **GLUT (glucose transporter) proteins** that transport glucose ([\[link\]](#)). Different forms of the GLUT protein control passage of glucose into the cells of specific tissues.



GLUT4 is a glucose transporter that is stored in vesicles. A cascade of events that occurs upon insulin binding to a receptor in the plasma membrane causes GLUT4-containing vesicles to fuse with the plasma membrane so that glucose may be transported into the cell.

Some reactions are controlled by having two different enzymes—one each for the two directions of a reversible reaction. Reactions that are catalyzed by only one enzyme can go to equilibrium, stalling the reaction. In contrast, if two different enzymes (each specific for a given direction) are necessary for a reversible reaction, the opportunity to control the rate of the reaction increases, and equilibrium is not reached.

A number of enzymes involved in each of the pathways—in particular, the enzyme catalyzing the first committed reaction of the pathway—are controlled by attachment of a molecule to an allosteric site on the protein. The molecules most commonly used in this capacity are the nucleotides ATP, ADP, AMP, NAD^+ , and NADH. These regulators—allosteric effectors—may increase or decrease enzyme activity, depending on the prevailing conditions. The allosteric effector alters the steric structure of the enzyme,

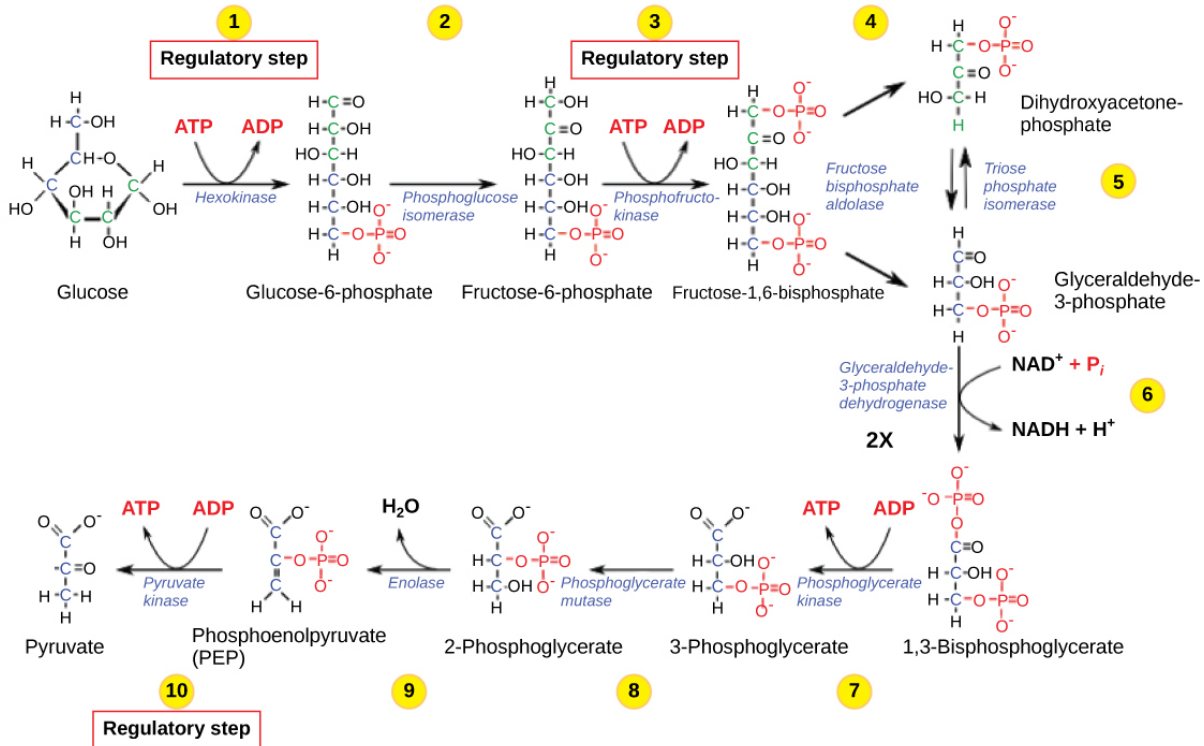
usually affecting the configuration of the active site. This alteration of the protein's (the enzyme's) structure either increases or decreases its affinity for its substrate, with the effect of increasing or decreasing the rate of the reaction. The attachment signals to the enzyme. This binding can increase or decrease the enzyme's activity, providing a feedback mechanism. This feedback type of control is effective as long as the chemical affecting it is attached to the enzyme. Once the overall concentration of the chemical decreases, it will diffuse away from the protein, and the control is relaxed.

Control of Catabolic Pathways

Enzymes, proteins, electron carriers, and pumps that play roles in glycolysis, the citric acid cycle, and the electron transport chain tend to catalyze nonreversible reactions. In other words, if the initial reaction takes place, the pathway is committed to proceeding with the remaining reactions. Whether a particular enzyme activity is released depends upon the energy needs of the cell (as reflected by the levels of ATP, ADP, and AMP).

Glycolysis

The control of glycolysis begins with the first enzyme in the pathway, hexokinase ([\[link\]](#)). This enzyme catalyzes the phosphorylation of glucose, which helps to prepare the compound for cleavage in a later step. The presence of the negatively charged phosphate in the molecule also prevents the sugar from leaving the cell. When hexokinase is inhibited, glucose diffuses out of the cell and does not become a substrate for the respiration pathways in that tissue. The product of the hexokinase reaction is glucose-6-phosphate, which accumulates when a later enzyme, phosphofructokinase, is inhibited.



The glycolysis pathway is primarily regulated at the three key enzymatic steps (1, 3, and 10) as indicated. Note that the first two steps that are regulated occur early in the pathway and involve hydrolysis of ATP.

Phosphofructokinase is the main enzyme controlled in glycolysis. High levels of ATP or citrate or a lower, more acidic pH decreases the enzyme's activity. An increase in citrate concentration can occur because of a blockage in the citric acid cycle. Fermentation, with its production of organic acids such as lactic acid, frequently accounts for the increased acidity in a cell; however, the products of fermentation do not typically accumulate in cells.

The last step in glycolysis is catalyzed by pyruvate kinase. The pyruvate produced can proceed to be catabolized or converted into the amino acid alanine. If no more energy is needed and alanine is in adequate supply, the enzyme is inhibited. The enzyme's activity is increased when fructose-1,6-bisphosphate levels increase. (Recall that fructose-1,6-bisphosphate is an

intermediate in the first half of glycolysis.) The regulation of pyruvate kinase involves phosphorylation by a kinase (pyruvate kinase), resulting in a less-active enzyme. Dephosphorylation by a phosphatase reactivates it. Pyruvate kinase is also regulated by ATP (a negative allosteric effect).

If more energy is needed, more pyruvate will be converted into acetyl CoA through the action of pyruvate dehydrogenase. If either acetyl groups or NADH accumulates, there is less need for the reaction, and the rate decreases. Pyruvate dehydrogenase is also regulated by phosphorylation: a kinase phosphorylates it to form an inactive enzyme, and a phosphatase reactivates it. The kinase and the phosphatase are also regulated.

Citric Acid Cycle

The citric acid cycle is controlled through the enzymes that catalyze the reactions that make the first two molecules of NADH ([\[link\]](#)). These enzymes are isocitrate dehydrogenase and α -ketoglutarate dehydrogenase. When adequate ATP and NADH levels are available, the rates of these reactions decrease. When more ATP is needed, as reflected in rising ADP levels, the rate increases. Alpha-ketoglutarate dehydrogenase will also be affected by the levels of succinyl CoA—a subsequent intermediate in the cycle—causing a decrease in activity. A decrease in the rate of operation of the pathway at this point is not necessarily negative, as the increased levels of the α -ketoglutarate not used by the citric acid cycle can be used by the cell for amino acid (glutamate) synthesis.

Electron Transport Chain

Specific enzymes of the electron transport chain are unaffected by feedback inhibition, but the rate of electron transport through the pathway is affected by the levels of ADP and ATP. Greater ATP consumption by a cell is indicated by a buildup of ADP. As ATP usage decreases, the concentration of ADP decreases, and now, ATP begins to build up in the cell. This change

in the relative concentration of ADP to ATP triggers the cell to slow down the electron transport chain.

Note:

Link to Learning

Visit this [site](#) to see an animation of the electron transport chain and ATP synthesis.

For a summary of feedback controls in cellular respiration, see [\[link\]](#).

Summary of Feedback Controls in Cellular Respiration			
Pathway	Enzyme affected	Elevated levels of effector	Effect on pathway activity
glycolysis	hexokinase	glucose-6-phosphate	decrease
	phosphofructokinase	low-energy charge (ATP, AMP), fructose-6-phosphate via fructose-2,6-bisphosphate	increase

Summary of Feedback Controls in Cellular Respiration			
Pathway	Enzyme affected	Elevated levels of effector	Effect on pathway activity
		high-energy charge (ATP, AMP), citrate, acidic pH	decrease
	pyruvate kinase	fructose-1,6-bisphosphate	increase
		high-energy charge (ATP, AMP), alanine	decrease
pyruvate to acetyl CoA conversion	pyruvate dehydrogenase	ADP, pyruvate	increase
		acetyl CoA, ATP, NADH	decrease
citric acid cycle	isocitrate dehydrogenase	ADP	increase
		ATP, NADH	decrease
	α -ketoglutarate dehydrogenase	calcium ions, ADP	increase

Summary of Feedback Controls in Cellular Respiration			
Pathway	Enzyme affected	Elevated levels of effector	Effect on pathway activity
		ATP, NADH, succinyl CoA	decrease
electron transport chain		ADP	increase
		ATP	decrease

Section Summary

Cellular respiration is controlled by a variety of means. The entry of glucose into a cell is controlled by the transport proteins that aid glucose passage through the cell membrane. Most of the control of the respiration processes is accomplished through the control of specific enzymes in the pathways. This is a type of negative feedback mechanism, turning the enzymes off. The enzymes respond most often to the levels of the available nucleosides ATP, ADP, AMP, NAD^+ , and FAD. Other intermediates of the pathway also affect certain enzymes in the systems.

Review Questions

Exercise:

Problem:

The effect of high levels of ADP is to _____ in cellular respiration.

- a. increase the activity of specific enzymes
 - b. decrease the activity of specific enzymes
 - c. have no effect on the activity of specific enzymes
 - d. slow down the pathway
-

Solution:

A

Exercise:

Problem:

The control of which enzyme exerts the most control on glycolysis?

- a. hexokinase
 - b. phosphofructokinase
 - c. glucose-6-phosphatase
 - d. aldolase
-

Solution:

B

Critical Thinking Questions

Exercise:

Problem:How does citrate from the citric acid cycle affect glycolysis?

Solution:

Citrate can inhibit phosphofructokinase by feedback regulation.

Exercise:

Problem:

Why might negative feedback mechanisms be more common than positive feedback mechanisms in living cells?

Solution:

Negative feedback mechanisms actually control a process; it can turn it off, whereas positive feedback accelerates the process, allowing the cell no control over it. Negative feedback naturally maintains homeostasis, whereas positive feedback drives the system away from equilibrium.

Glossary

GLUT protein

integral membrane protein that transports glucose

Overview of Photosynthesis

By the end of this section, you will be able to do the following:

- Explain the significance of photosynthesis to other living organisms
- Describe the main structures involved in photosynthesis
- Identify the substrates and products of photosynthesis

Photosynthesis is essential to all life on earth; both plants and animals depend on it. It is the only biological process that can capture energy that originates from sunlight and converts it into chemical compounds (carbohydrates) that every organism uses to power its metabolism. It is also a source of oxygen necessary for many living organisms. In brief, the energy of sunlight is “captured” to energize electrons, whose energy is then stored in the covalent bonds of sugar molecules. How long lasting and stable are those covalent bonds? The energy extracted today by the burning of coal and petroleum products represents sunlight energy captured and stored by photosynthesis 350 to 200 million years ago during the Carboniferous Period.

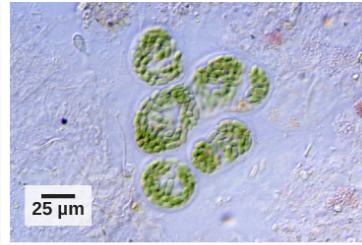
Plants, algae, and a group of bacteria called cyanobacteria are the only organisms capable of performing photosynthesis ([\[link\]](#)). Because they use light to manufacture their own food, they are called **photoautotrophs** (literally, “self-feeders using light”). Other organisms, such as animals, fungi, and most other bacteria, are termed **heterotrophs** (“other feeders”), because they must rely on the sugars produced by photosynthetic organisms for their energy needs. A third very interesting group of bacteria synthesize sugars, not by using sunlight’s energy, but by extracting energy from inorganic chemical compounds. For this reason, they are referred to as **chemoautotrophs**.



(a)



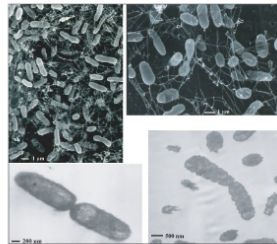
(b)



(c)



(d)



(e)

Photoautotrophs including (a) plants, (b) algae, and (c) cyanobacteria synthesize their organic compounds via photosynthesis using sunlight as an energy source. Cyanobacteria and planktonic algae can grow over enormous areas in water, at times completely covering the surface. In a (d) deep sea vent, chemoautotrophs, such as these (e) thermophilic bacteria, capture energy from inorganic compounds to produce organic compounds. The ecosystem surrounding the vents has a diverse array of animals, such as tubeworms, crustaceans, and octopuses that derive energy from the bacteria. (credit a: modification of work by Steve Hillebrand, U.S. Fish and Wildlife Service; credit b: modification of work by "eutrophication&hypoxia"/Flickr; credit c: modification of work by NASA; credit d: University of Washington, NOAA; credit e: modification of work by Mark Amend, West Coast and Polar Regions Undersea Research Center, UAF, NOAA)

The importance of photosynthesis is not just that it can capture sunlight's energy. After all, a lizard sunning itself on a cold day can use the sun's energy to warm up in a process called *behavioral thermoregulation*. In

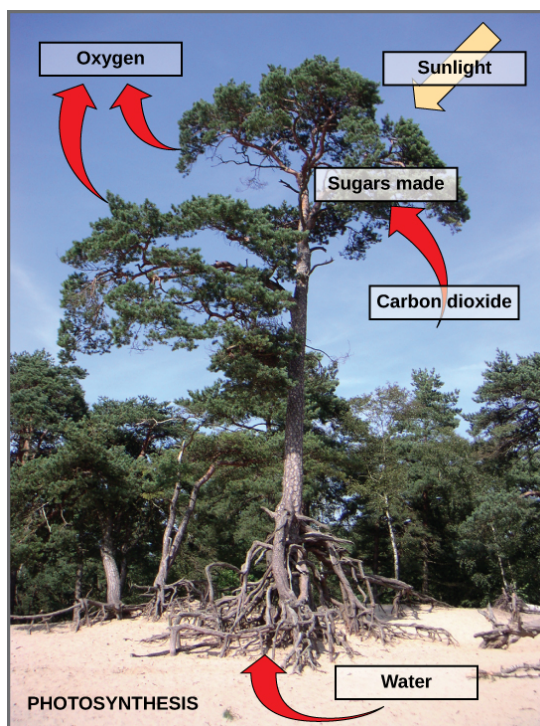
contrast, photosynthesis is vital because it evolved as a way to *store the energy from solar radiation (the “photo-” part) to energy in the carbon-carbon bonds of carbohydrate molecules (the “-synthesis” part)*. Those carbohydrates are the energy source that heterotrophs use to power the synthesis of ATP via respiration. Therefore, photosynthesis powers 99 percent of Earth’s ecosystems. When a top predator, such as a wolf, preys on a deer ([\[link\]](#)), the wolf is at the end of an energy path that went from nuclear reactions on the surface of the sun, to visible light, to photosynthesis, to vegetation, to deer, and finally to the wolf.



The energy stored in carbohydrate molecules from photosynthesis passes through the food chain. The predator that eats these deer receives a portion of the energy that originated in the photosynthetic vegetation that the deer consumed. (credit: modification of work by Steve VanRiper, U.S. Fish and Wildlife Service)

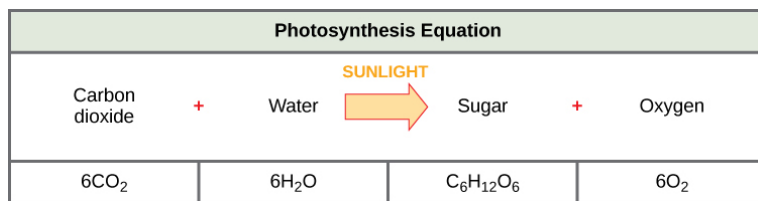
Main Structures and Summary of Photosynthesis

Photosynthesis is a multi-step process that requires specific wavelengths of visible sunlight, carbon dioxide (which is low in energy), and water as substrates ([\[link\]](#)). After the process is complete, it releases oxygen and produces glyceraldehyde-3-phosphate (GA3P), as well as simple carbohydrate molecules (high in energy) that can then be converted into glucose, sucrose, or any of dozens of other sugar molecules. These sugar molecules contain energy and the energized carbon that all living things need to survive.



Photosynthesis uses solar energy, carbon dioxide, and water to produce energy-storing carbohydrates. Oxygen is generated as a waste product of photosynthesis.

The following is the chemical equation for photosynthesis ([link](#)):



The basic equation for photosynthesis is deceptively simple. In reality, the process takes place in many steps involving intermediate reactants and products. Glucose, the primary energy source in cells, is made from two three-carbon GA3Ps.

Although the equation looks simple, the many steps that take place during photosynthesis are actually quite complex. Before learning the details of how photoautotrophs turn sunlight into food, it is important to become familiar with the structures involved.

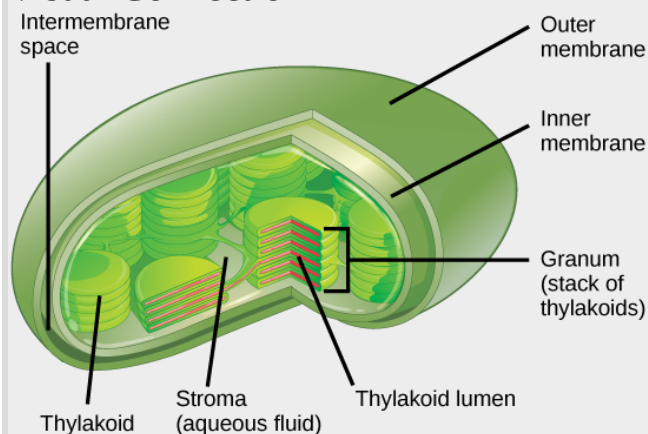
Basic Photosynthetic Structures

In plants, photosynthesis generally takes place in leaves, which consist of several layers of cells. The process of photosynthesis occurs in a middle layer called the **mesophyll**. The gas exchange of carbon dioxide and oxygen occurs through small, regulated openings called **stomata** (singular: stoma), which also play roles in the regulation of gas exchange and water balance. The stomata are typically located on the underside of the leaf, which helps to minimize water loss due to high temperatures on the upper surface of the leaf. Each stoma is flanked by guard cells that regulate the opening and closing of the stomata by swelling or shrinking in response to osmotic changes.

In all autotrophic eukaryotes, photosynthesis takes place inside an organelle called a **chloroplast**. For plants, chloroplast-containing cells exist mostly in the mesophyll. Chloroplasts have a double membrane envelope (composed of an outer membrane and an inner membrane), and are ancestrally derived from ancient free-living cyanobacteria. Within the chloroplast are stacked, disc-shaped structures called **thylakoids**. Embedded in the thylakoid membrane is chlorophyll, a **pigment** (molecule that absorbs light) responsible for the initial interaction between light and plant material, and numerous proteins that make up the electron transport chain. The thylakoid membrane encloses an internal space called the **thylakoid lumen**. As shown in [\[link\]](#), a stack of thylakoids is called a **granum**, and the liquid-filled space surrounding the granum is called **stroma** or “bed” (not to be confused with stoma or “mouth,” an opening on the leaf epidermis).

Note:

Visual Connection

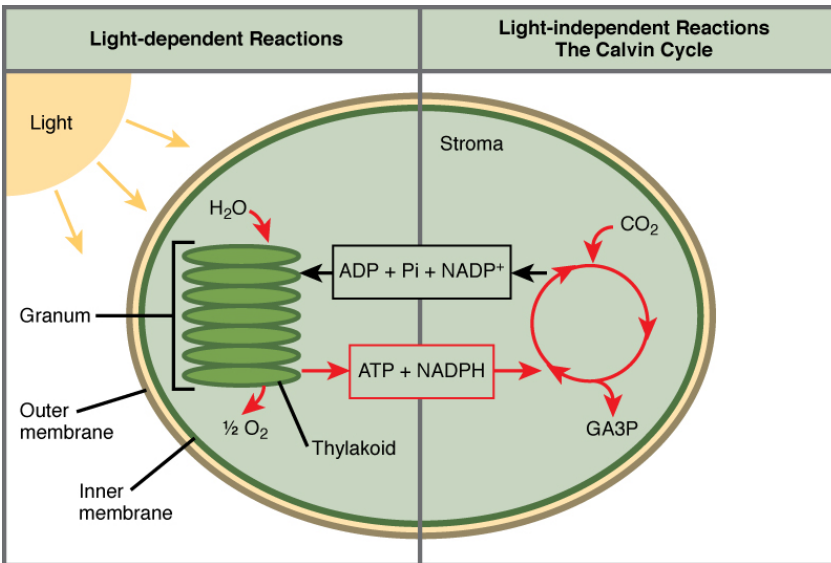


Photosynthesis takes place in chloroplasts, which have an outer membrane and an inner membrane. Stacks of thylakoids called grana form a third membrane layer.

On a hot, dry day, the guard cells of plants close their stomata to conserve water. What impact will this have on photosynthesis?

The Two Parts of Photosynthesis

Photosynthesis takes place in two sequential stages: the light-dependent reactions and the light-independent reactions. In the **light-dependent reactions**, energy from sunlight is absorbed by chlorophyll and that energy is converted into stored chemical energy. In the **light-independent reactions**, the chemical energy harvested during the light-dependent reactions drives the assembly of sugar molecules from carbon dioxide. Therefore, although the light-independent reactions do not use light as a reactant, they require the products of the light-dependent reactions to function. In addition, however, several enzymes of the light-independent reactions are activated by light. The light-dependent reactions utilize certain molecules to temporarily store the energy: These are referred to as *energy carriers*. The energy carriers that move energy from light-dependent reactions to light-independent reactions can be thought of as “full” because they are rich in energy. After the energy is released, the “empty” energy carriers return to the light-dependent reaction to obtain more energy. [\[link\]](#) illustrates the components inside the chloroplast where the light-dependent and light-independent reactions take place.



Photosynthesis takes place in two stages: light-dependent reactions and the Calvin cycle. Light-dependent reactions, which take place in the thylakoid membrane, use light energy to make ATP and NADPH. The Calvin cycle, which takes place in the stroma, uses energy derived from these compounds to make GA3P from CO_2 .

Note:

Link to Learning

Click the [link](#) to learn more about photosynthesis.

Note:

Everyday Connection

Photosynthesis at the Grocery Store



Foods that humans consume originate from photosynthesis.
(credit: Associação Brasileira de Supermercados)

Major grocery stores in the United States are organized into departments, such as dairy, meats, produce, bread, cereals, and so forth. Each aisle ([link](#)) contains hundreds, if not thousands, of different products for customers to buy and consume.

Although there is a large variety, each item ultimately can be linked back to photosynthesis. Meats and dairy link, because the animals were fed plant-based foods. The breads, cereals, and pastas come largely from starchy grains, which are the seeds of photosynthesis-dependent plants. What about desserts and drinks? All of these products contain sugar—sucrose is a plant product, a disaccharide, a carbohydrate molecule, which is built directly from photosynthesis. Moreover, many items are less obviously derived from plants: For instance, paper goods are generally plant products, and many plastics (abundant as products and packaging) are derived from “algae” (unicellular plant-like organisms, and cyanobacteria). Virtually every spice and flavoring in the spice aisle was produced by a plant as a leaf, root, bark, flower, fruit, or stem. Ultimately, photosynthesis connects to every meal and every food a person consumes.

Section Summary

The process of photosynthesis transformed life on Earth. By harnessing energy from the sun, the evolution of photosynthesis allowed living things access to enormous amounts of energy. Because of photosynthesis, living things gained access to sufficient energy that allowed them to build new structures and achieve the biodiversity evident today.

Only certain organisms (photoautotrophs), can perform photosynthesis; they require the presence of chlorophyll, a specialized pigment that absorbs certain wavelengths of the visible spectrum and can capture energy from sunlight. Photosynthesis uses carbon dioxide and water to assemble carbohydrate molecules and release oxygen as a byproduct into the atmosphere. Eukaryotic autotrophs, such as plants and algae, have organelles called chloroplasts in which photosynthesis takes place, and starch accumulates. In prokaryotes, such as cyanobacteria, the process is less localized and occurs within folded membranes, extensions of the plasma membrane, and in the cytoplasm.

Visual Connection Questions

Exercise:

Problem:

[\[link\]](#) On a hot, dry day, the guard cells of plants close their stomata to conserve water. What impact will this have on photosynthesis?

Solution:

[\[link\]](#) Levels of carbon dioxide (a necessary photosynthetic substrate) will immediately fall. As a result, the rate of photosynthesis will be inhibited.

Review Questions

Exercise:

Problem:

Which of the following components is *not* used by both plants and cyanobacteria to carry out photosynthesis?

- a. chloroplasts
- b. chlorophyll
- c. carbon dioxide
- d. water

Solution:

A

Exercise:

Problem: What two main products result from photosynthesis?

- a. oxygen and carbon dioxide
- b. chlorophyll and oxygen
- c. sugars/carbohydrates and oxygen
- d. sugars/carbohydrates and carbon dioxide

Solution:

C

Exercise:

Problem:

In which compartment of the plant cell do the light-independent reactions of photosynthesis take place?

- a. thylakoid
- b. stroma
- c. outer membrane

d. mesophyll

Solution:

B

Exercise:

Problem:

Which statement about thylakoids in eukaryotes is *not* correct?

- a. Thylakoids are assembled into stacks.
- b. Thylakoids exist as a maze of folded membranes.
- c. The space surrounding thylakoids is called stroma.
- d. Thylakoids contain chlorophyll.

Solution:

B

Exercise:

Problem:

Predict the end result if a chloroplast's light-independent enzymes developed a mutation that prevented them from activating in response to light.

- a. GA3P accumulation
- b. ATP and NADPH accumulation
- c. Water accumulation
- d. Carbon dioxide depletion

Solution:

B

Exercise:**Problem:**

How are the NADPH and GA3P molecules made during photosynthesis similar?

- a. They are both end products of photosynthesis.
- b. They are both substrates for photosynthesis.
- c. They are both produced from carbon dioxide.
- d. They both store energy in chemical bonds.

Solution:

D

Critical Thinking Questions**Exercise:****Problem:**

What is the overall outcome of the light reactions in photosynthesis?

Solution:

The outcome of light reactions in photosynthesis is the conversion of solar energy into chemical energy that the chloroplasts can use to do work (mostly anabolic production of carbohydrates from carbon dioxide).

Exercise:**Problem:**

Why are carnivores, such as lions, dependent on photosynthesis to survive?

Solution:

Because lions eat animals that eat plants.

Exercise:**Problem:**

Why are energy carriers thought of as either “full” or “empty”?

Solution:

The energy carriers that move from the light-dependent reaction to the light-independent one are “full” because they bring energy. After the energy is released, the “empty” energy carriers return to the light-dependent reaction to obtain more energy. There is not much actual movement involved. Both ATP and NADPH are produced in the stroma where they are also used and reconverted into ADP, Pi, and NADP⁺.

Exercise:**Problem:**

Describe how the grey wolf population would be impacted by a volcanic eruption that spewed a dense ash cloud that blocked sunlight in a section of Yellowstone National Park.

Solution:

The grey wolves are apex predators in their food web, meaning they consume smaller prey animals and are not the prey of any other animal. Blocking sunlight would prevent the plants at the bottom of the food web from performing photosynthesis. This would kill many of the plants, reducing the food sources available to smaller animals in Yellowstone. A smaller prey animal population means that fewer wolves can survive in the area, and the population of grey wolves will decrease.

Exercise:

Problem: How does the closing of the stomata limit photosynthesis?

Solution:

The stomata regulate the exchange of gases and water vapor between a leaf and its surrounding environment. When the stomata are closed, the water molecules cannot escape the leaf, but the leaf also cannot acquire new carbon dioxide molecules from the environment. This limits the light-independent reactions to only continuing until the carbon dioxide stores in the leaf are depleted.

Glossary

chemoautotroph

organism that can build organic molecules using energy derived from inorganic chemicals instead of sunlight

chloroplast

organelle in which photosynthesis takes place

granum

stack of thylakoids located inside a chloroplast

heterotroph

organism that consumes organic substances or other organisms for food

light-dependent reaction

first stage of photosynthesis where certain wavelengths of the visible light are absorbed to form two energy-carrying molecules (ATP and NADPH)

light-independent reaction

second stage of photosynthesis, through which carbon dioxide is used to build carbohydrate molecules using energy from ATP and NADPH

mesophyll

middle layer of chlorophyll-rich cells in a leaf

photoautotroph

organism capable of producing its own organic compounds from sunlight

pigment

molecule that is capable of absorbing certain wavelengths of light and reflecting others (which accounts for its color)

stoma

opening that regulates gas exchange and water evaporation between leaves and the environment, typically situated on the underside of leaves

stroma

fluid-filled space surrounding the grana inside a chloroplast where the light-independent reactions of photosynthesis take place

thylakoid

disc-shaped, membrane-bound structure inside a chloroplast where the light-dependent reactions of photosynthesis take place; stacks of thylakoids are called grana

thylakoid lumen

aqueous space bound by a thylakoid membrane where protons accumulate during light-driven electron transport

The Light-Dependent Reactions of Photosynthesis

By the end of this section, you will be able to do the following:

- Explain how plants absorb energy from sunlight
- Describe short and long wavelengths of light
- Describe how and where photosynthesis takes place within a plant

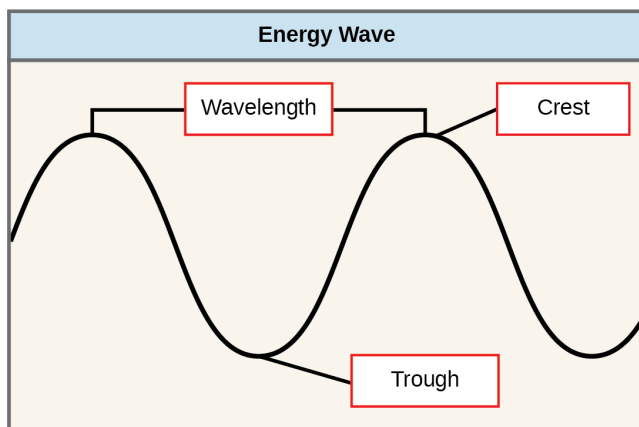
How can light energy be used to make food? When a person turns on a lamp, electrical energy becomes light energy. Like all other forms of kinetic energy, light can travel, change form, and be harnessed to do work. In the case of photosynthesis, light energy is converted into chemical energy, which photoautotrophs use to build basic carbohydrate molecules ([\[link\]](#)). However, autotrophs only use a few specific wavelengths of sunlight.



Photoautotrophs can capture visible light energy in specific wavelengths from the sun, converting it into the chemical energy used to build food molecules. (credit: Gerry Atwell)

What Is Light Energy?

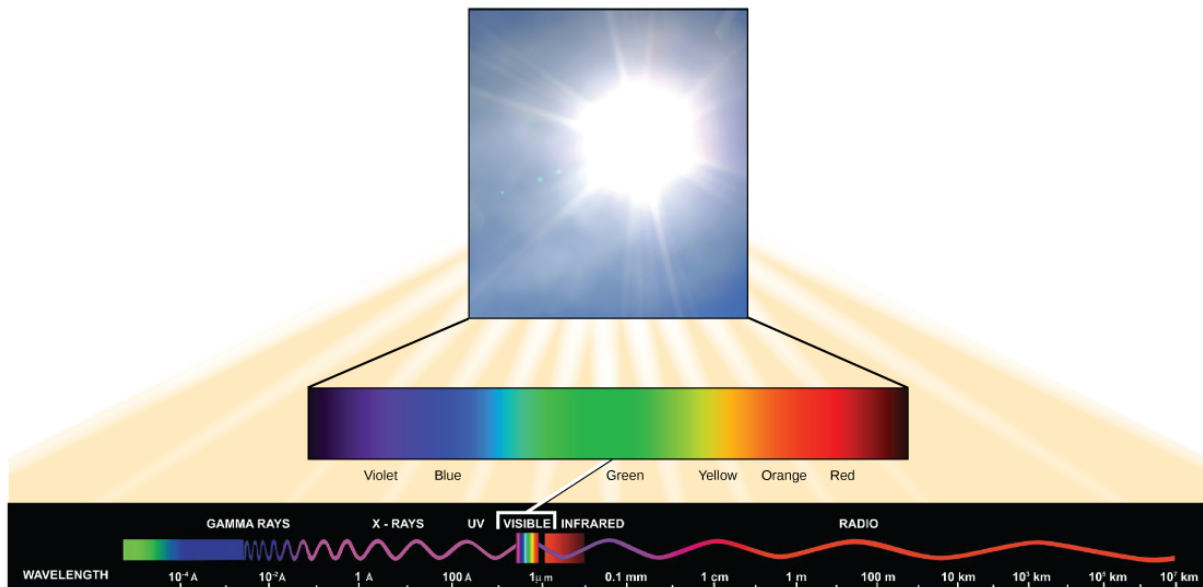
The sun emits an enormous amount of electromagnetic radiation (solar energy in a spectrum from very short gamma rays to very long radio waves). Humans can see only a tiny fraction of this energy, which we refer to as “visible light.” The manner in which solar energy travels is described as waves. Scientists can determine the amount of energy of a wave by measuring its **wavelength** (shorter wavelengths are more powerful than longer wavelengths)—the distance between consecutive crest points of a wave. Therefore, a single wave is measured from two consecutive points, such as from crest to crest or from trough to trough ([\[link\]](#)).



The wavelength of a single wave is the distance between two consecutive points of similar position (two crests or two troughs) along the wave.

Visible light constitutes only one of many types of electromagnetic radiation emitted from the sun and other stars. Scientists differentiate the

various types of radiant energy from the sun within the electromagnetic spectrum. The **electromagnetic spectrum** is the range of all possible frequencies of radiation ([\[link\]](#)). The difference between wavelengths relates to the amount of energy carried by them.



The sun emits energy in the form of electromagnetic radiation. This radiation exists at different wavelengths, each of which has its own characteristic energy. All electromagnetic radiation, including visible light, is characterized by its wavelength.

Each type of electromagnetic radiation travels at a particular wavelength. The longer the wavelength, the less energy it carries. Short, tight waves carry the most energy. This may seem illogical, but think of it in terms of a piece of moving heavy rope. It takes little effort by a person to move a rope in long, wide waves. To make a rope move in short, tight waves, a person would need to apply significantly more energy.

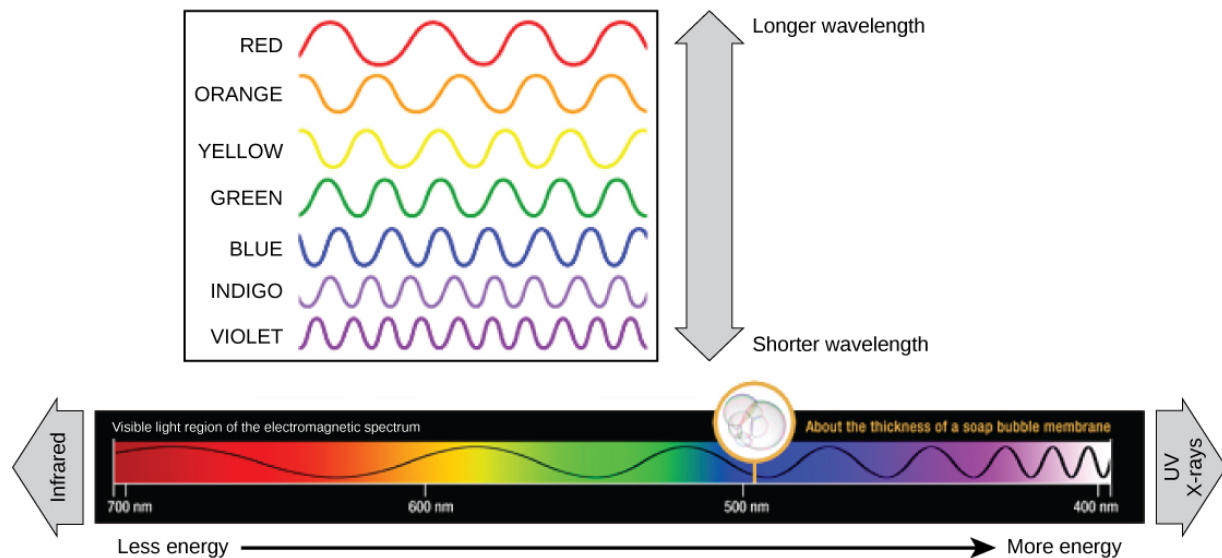
The electromagnetic spectrum ([\[link\]](#)) shows several types of electromagnetic radiation originating from the sun, including X-rays and ultraviolet (UV) rays. The higher-energy waves can penetrate tissues and

damage cells and DNA, which explains why both X-rays and UV rays can be harmful to living organisms.

Absorption of Light

Light energy initiates the process of photosynthesis when pigments absorb specific wavelengths of visible light. Organic pigments, whether in the human retina or the chloroplast thylakoid, have a narrow range of energy levels that they can absorb. Energy levels lower than those represented by red light are insufficient to raise an orbital electron to an excited (quantum) state. Energy levels higher than those in blue light will physically tear the molecules apart, in a process called bleaching. Our retinal pigments can only “see” (absorb) wavelengths between 700 nm and 400 nm of light, a spectrum that is therefore called visible light. For the same reasons, plants, pigment molecules absorb only light in the wavelength range of 700 nm to 400 nm; plant physiologists refer to this range for plants as photosynthetically active radiation.

The visible light seen by humans as white light actually exists in a rainbow of colors. Certain objects, such as a prism or a drop of water, disperse white light to reveal the colors to the human eye. The visible light portion of the electromagnetic spectrum shows the rainbow of colors, with violet and blue having shorter wavelengths, and therefore higher energy. At the other end of the spectrum toward red, the wavelengths are longer and have lower energy ([\[link\]](#)).



The colors of visible light do not carry the same amount of energy. Violet has the shortest wavelength and therefore carries the most energy, whereas red has the longest wavelength and carries the least amount of energy. (credit: modification of work by NASA)

Understanding Pigments

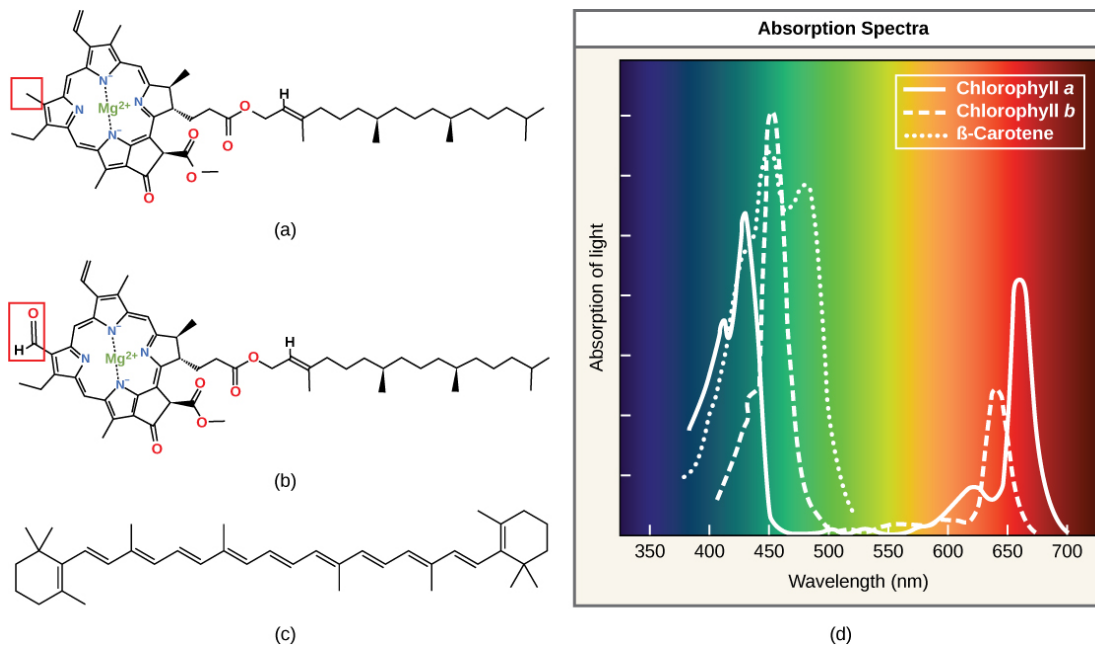
Different kinds of pigments exist, and each absorbs only specific wavelengths (colors) of visible light. Pigments reflect or transmit the wavelengths they cannot absorb, making them appear a mixture of the reflected or transmitted light colors.

Chlorophylls and carotenoids are the two major classes of photosynthetic pigments found in plants and algae; each class has multiple types of pigment molecules. There are five major chlorophylls: *a*, *b*, *c* and *d* and a related molecule found in prokaryotes called *bacteriochlorophyll*.

Chlorophyll *a* and **chlorophyll *b*** are found in higher plant chloroplasts and will be the focus of the following discussion.

With dozens of different forms, carotenoids are a much larger group of pigments. The carotenoids found in fruit—such as the red of tomato (lycopene), the yellow of corn seeds (zeaxanthin), or the orange of an orange peel (β -carotene)—are used as advertisements to attract seed dispersers. In photosynthesis, **carotenoids** function as photosynthetic pigments that are very efficient molecules for the disposal of excess energy. When a leaf is exposed to full sun, the light-dependent reactions are required to process an enormous amount of energy; if that energy is not handled properly, it can do significant damage. Therefore, many carotenoids reside in the thylakoid membrane, absorb excess energy, and safely dissipate that energy as heat.

Each type of pigment can be identified by the specific pattern of wavelengths it absorbs from visible light: This is termed the **absorption spectrum**. The graph in [\[link\]](#) shows the absorption spectra for chlorophyll *a*, chlorophyll *b*, and a type of carotenoid pigment called β -carotene (which absorbs blue and green light). Notice how each pigment has a distinct set of peaks and troughs, revealing a highly specific pattern of absorption. Chlorophyll *a* absorbs wavelengths from either end of the visible spectrum (blue and red), but not green. Because green is reflected or transmitted, chlorophyll appears green. Carotenoids absorb in the short-wavelength blue region, and reflect the longer yellow, red, and orange wavelengths.



(a) Chlorophyll *a*, (b) chlorophyll *b*, and (c) β -carotene are hydrophobic organic pigments found in the thylakoid membrane. Chlorophyll *a* and *b*, which are identical except for the part indicated in the red box, are responsible for the green color of leaves. β -carotene is responsible for the orange color in carrots. Each pigment has (d) a unique absorbance spectrum.

Many photosynthetic organisms have a mixture of pigments, and by using these pigments, the organism can absorb energy from a wider range of wavelengths. Not all photosynthetic organisms have full access to sunlight. Some organisms grow underwater where light intensity and quality decrease and change with depth. Other organisms grow in competition for light. Plants on the rainforest floor must be able to absorb any bit of light that comes through, because the taller trees absorb most of the sunlight and scatter the remaining solar radiation ([\[link\]](#)).



Plants that commonly grow in the shade have adapted to low levels of light by changing the relative concentrations of their chlorophyll pigments. (credit: Jason Hollinger)

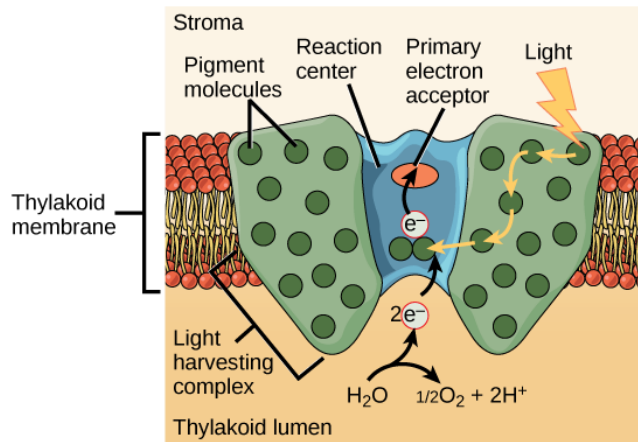
When studying a photosynthetic organism, scientists can determine the types of pigments present by generating absorption spectra. An instrument called a **spectrophotometer** can differentiate which wavelengths of light a substance can absorb. Spectrophotometers measure transmitted light and compute from it the absorption. By extracting pigments from leaves and placing these samples into a spectrophotometer, scientists can identify which wavelengths of light an organism can absorb. Additional methods for the identification of plant pigments include various types of chromatography that separate the pigments by their relative affinities to solid and mobile phases.

How Light-Dependent Reactions Work

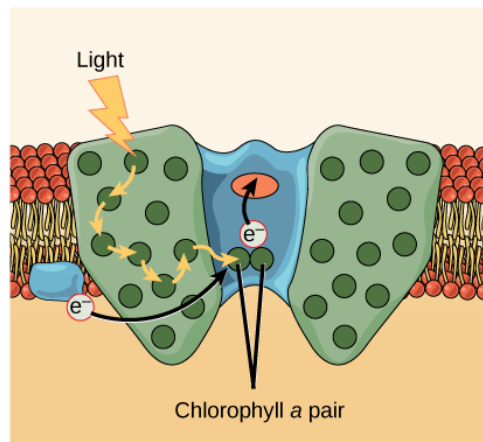
The overall function of light-dependent reactions is to convert solar energy into chemical energy in the form of NADPH and ATP. This chemical energy supports the light-independent reactions and fuels the assembly of sugar molecules. The light-dependent reactions are depicted in [\[link\]](#).

Protein complexes and pigment molecules work together to produce NADPH and ATP. The numbering of the photosystems is derived from the order in which they were discovered, not in the order of the transfer of electrons.

(a) Photosystem II (P680)



(b) Photosystem I (P700)

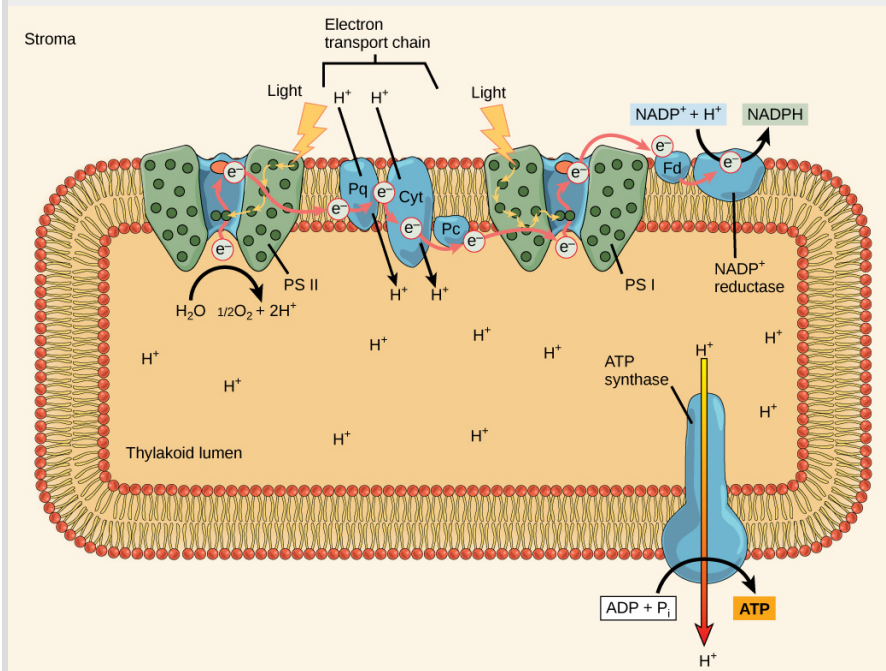


A photosystem consists of 1) a light-harvesting complex and 2) a reaction center. Pigments in the light-harvesting complex pass light energy to two special chlorophyll *a* molecules in the reaction center. The light excites

an electron from the chlorophyll *a* pair, which passes to the primary electron acceptor. The excited electron must then be replaced. In (a) photosystem II, the electron comes from the splitting of water, which releases oxygen as a waste product. In (b) photosystem I, the electron comes from the chloroplast electron transport chain discussed below.

The actual step that converts light energy into chemical energy takes place in a multiprotein complex called a **photosystem**, two types of which are found embedded in the thylakoid membrane: **photosystem II** (PSII) and **photosystem I** (PSI) ([\[link\]](#)). The two complexes differ on the basis of what they oxidize (that is, the source of the low-energy electron supply) and what they reduce (the place to which they deliver their energized electrons).

Both photosystems have the same basic structure; a number of **antenna proteins** to which the chlorophyll molecules are bound surround the **reaction center** where the photochemistry takes place. Each photosystem is serviced by the **light-harvesting complex**, which passes energy from sunlight to the reaction center; it consists of multiple antenna proteins that contain a mixture of 300 to 400 chlorophyll *a* and *b* molecules as well as other pigments like carotenoids. The absorption of a single **photon** or distinct quantity or “packet” of light by any of the chlorophylls pushes that molecule into an excited state. In short, the light energy has now been captured by biological molecules but is not stored in any useful form yet. The energy is transferred from chlorophyll to chlorophyll until eventually (after about a millionth of a second), it is delivered to the reaction center. Up to this point, only energy has been transferred between molecules, not electrons.

Note:**Visual Connection**

In the photosystem II (PSII) reaction center, energy from sunlight is used to extract electrons from water. The electrons travel through the chloroplast electron transport chain to photosystem I (PSI), which reduces NADP^+ to NADPH. The electron transport chain moves protons across the thylakoid membrane into the lumen. At the same time, splitting of water adds protons to the lumen, and reduction of NADPH removes protons from the stroma. The net result is a low pH in the thylakoid lumen, and a high pH in the stroma. ATP synthase uses this electrochemical gradient to make ATP.

What is the initial source of electrons for the chloroplast electron transport chain?

a. water

- b. oxygen
- c. carbon dioxide
- d. NADPH

The reaction center contains a pair of chlorophyll *a* molecules with a special property. Those two chlorophylls can undergo oxidation upon excitation; they can actually give up an electron in a process called a **photoact**. It is at this step in the reaction center during photosynthesis that light energy is converted into an excited electron. All of the subsequent steps involve getting that electron onto the energy carrier NADPH for delivery to the Calvin cycle where the electron is deposited onto carbon for long-term storage in the form of a carbohydrate. PSII and PSI are two major components of the photosynthetic **electron transport chain**, which also includes the *cytochrome complex*. The cytochrome complex, an enzyme composed of two protein complexes, transfers the electrons from the carrier molecule plastoquinone (Pq) to the protein plastocyanin (Pc), thus enabling both the transfer of protons across the thylakoid membrane and the transfer of electrons from PSII to PSI.

The reaction center of PSII (called **P680**) delivers its high-energy electrons, one at the time, to the **primary electron acceptor**, and through the electron transport chain (Pq to cytochrome complex to plastocyanine) to PSI. P680's missing electron is replaced by extracting a low-energy electron from water; thus, water is "split" during this stage of photosynthesis, and PSII is re-reduced after every photoact. Splitting one H₂O molecule releases two electrons, two hydrogen atoms, and one atom of oxygen. However, splitting two molecules is required to form one molecule of diatomic O₂ gas. About 10 percent of the oxygen is used by mitochondria in the leaf to support oxidative phosphorylation. The remainder escapes to the atmosphere where it is used by aerobic organisms to support respiration.

As electrons move through the proteins that reside between PSII and PSI, they lose energy. This energy is used to move hydrogen atoms from the stromal side of the membrane to the thylakoid lumen. Those hydrogen atoms, plus the ones produced by splitting water, accumulate in the

thylakoid lumen and will be used to synthesize ATP in a later step. Because the electrons have lost energy prior to their arrival at PSI, they must be re-energized by PSI, hence, another photon is absorbed by the PSI antenna. That energy is relayed to the PSI reaction center (called **P700**). P700 is oxidized and sends a high-energy electron to NADP^+ to form NADPH. Thus, PSII captures the energy to create proton gradients to make ATP, and PSI captures the energy to reduce NADP^+ into NADPH. The two photosystems work in concert, in part, to guarantee that the production of NADPH will roughly equal the production of ATP. Other mechanisms exist to fine-tune that ratio to exactly match the chloroplast's constantly changing energy needs.

Generating an Energy Carrier: ATP

As in the intermembrane space of the mitochondria during cellular respiration, the buildup of hydrogen ions inside the thylakoid lumen creates a *concentration gradient*. The passive diffusion of hydrogen ions from high concentration (in the thylakoid lumen) to low concentration (in the stroma) is harnessed to create ATP, just as in the electron transport chain of cellular respiration. The ions build up energy because of diffusion and because they all have the same electrical charge, repelling each other.

To release this energy, hydrogen ions will rush through any opening, similar to water jetting through a hole in a dam. In the thylakoid, that opening is a passage through a specialized protein channel called the ATP synthase. The energy released by the hydrogen ion stream allows ATP synthase to attach a third phosphate group to ADP, which forms a molecule of ATP ([\[link\]](#)). The flow of hydrogen ions through ATP synthase is called chemiosmosis because the ions move from an area of high to an area of low concentration through a semi-permeable structure of the thylakoid.

Note:

Link to Learning

Visit this [site](#) and click through the animation to view the process of photosynthesis within a leaf.

Section Summary

The pigments of the first part of photosynthesis, the light-dependent reactions, absorb energy from sunlight. A photon strikes the antenna pigments of photosystem II to initiate photosynthesis. The energy travels to the reaction center that contains chlorophyll *a* and then to the electron transport chain, which pumps hydrogen ions into the thylakoid interior. This action builds up a high concentration of hydrogen ions. The hydrogen ions flow through ATP synthase during chemiosmosis to form molecules of ATP, which are used for the formation of sugar molecules in the second stage of photosynthesis. Photosystem I absorbs a second photon, which results in the formation of an NADPH molecule, another energy and reducing carrier for the light-independent reactions.

Visual Connection Questions

Exercise:

Problem:

[\[link\]](#) What is the source of electrons for the chloroplast electron transport chain?

- a. Water
- b. Oxygen
- c. Carbon dioxide
- d. NADPH

Solution:

[\[link\]](#) A.

Review Questions

Exercise:

Problem:

Which of the following structures is *not* a component of a photosystem?

- a. ATP synthase
- b. antenna molecule
- c. reaction center
- d. primary electron acceptor

Solution:

A

Exercise:

Problem:

How many photons does it take to fully reduce one molecule of NADP^+ to NADPH?

- a. 1
- b. 2
- c. 4
- d. 8

Solution:

B

Exercise:

Problem:

Which complex is *not* involved in the establishment of conditions for ATP synthesis?

- a. photosystem I
- b. ATP synthase
- c. photosystem II
- d. cytochrome complex

Solution:

C

Exercise:**Problem:**

From which component of the light-dependent reactions does NADPH form most directly?

- a. photosystem II
- b. photosystem I
- c. cytochrome complex
- d. ATP synthase

Solution:

B

Exercise:

Problem:

Three of the same species of plant are each grown under a different colored light for the same amount of time. Plant A is grown under blue light, Plant B is grown under green light, and Plant C is grown under orange light. Assuming the plants use only chlorophyll *a* and chlorophyll *b* for photosynthesis, what would be the predicted order of the plants from most growth to least growth?

- a. A, C, B
- b. A, B, C
- c. C, A, B
- d. B, A, C

Solution:

A

Exercise:**Problem:**

Plants containing only chlorophyll *b* are exposed to radiation with the following wavelengths: 10nm (x-rays), 450nm (blue light), 670nm (red light), and 800nm (infrared light). Which plants harness the most energy for photosynthesis?

- a. X-ray irradiated plants
- b. Blue light irradiated plants
- c. Red light irradiated plants
- d. Infrared irradiated plants

Solution:

B

Critical Thinking Questions

Exercise:

Problem:

Describe the pathway of electron transfer from photosystem II to photosystem I in light-dependent reactions.

Solution:

A photon of light hits an antenna molecule in photosystem II, and the energy released by it travels through other antenna molecules to the reaction center. The energy causes an electron to leave a molecule of chlorophyll *a* to a primary electron acceptor protein. The electron travels through the electron transport chain and is accepted by a pigment molecule in photosystem I.

Exercise:

Problem: What are the roles of ATP and NADPH in photosynthesis?

Solution:

Both of these molecules carry energy; in the case of NADPH, it has reducing power that is used to fuel the process of making carbohydrate molecules in light-independent reactions.

Exercise:

Problem:

How and why would the end products of photosynthesis be changed if a plant had a mutation that eliminated its photosystem II complex?

Solution:

Knocking out photosystem II would eliminate the production of oxygen and ATP during photosynthesis. Photosystem II splits water into oxygen atoms, hydrogen protons that remain in the thylakoid

lumen, and hydrogen-derived electrons that move from the reaction center into the electron transport chain. The transfer of an electron through the electron transport chain provides the energy to pump more protons into the thylakoid lumen to maintain a higher concentration of protons there. Moving protons across the thylakoid membrane back to the stroma provides the energy for ATP synthase to produce ATP. Without this proton gradient, ATP will not be synthesized.

Glossary

absorption spectrum

range of wavelengths of electromagnetic radiation absorbed by a given substance

antenna protein

pigment molecule that directly absorbs light and transfers the energy absorbed to other pigment molecules

carotenoid

photosynthetic pigment (yellow-orange-red) that functions to dispose of excess energy

chlorophyll *a*

form of chlorophyll that absorbs violet-blue and red light and consequently has a bluish-green color; the only pigment molecule that performs the photochemistry by getting excited and losing an electron to the electron transport chain

chlorophyll *b*

accessory pigment that absorbs blue and red-orange light and consequently has a yellowish-green tint

cytochrome complex

group of reversibly oxidizable and reducible proteins that forms part of the electron transport chain between photosystem II and photosystem I

electromagnetic spectrum

range of all possible frequencies of radiation

electron transport chain

group of proteins between PSII and PSI that pass energized electrons and use the energy released by the electrons to move hydrogen ions against their concentration gradient into the thylakoid lumen

light harvesting complex

complex that passes energy from sunlight to the reaction center in each photosystem; it consists of multiple antenna proteins that contain a mixture of 300 to 400 chlorophyll *a* and *b* molecules as well as other pigments like carotenoids

P680

reaction center of photosystem II

P700

reaction center of photosystem I

photoact

ejection of an electron from a reaction center using the energy of an absorbed photon

photon

distinct quantity or “packet” of light energy

photosystem

group of proteins, chlorophyll, and other pigments that are used in the light-dependent reactions of photosynthesis to absorb light energy and convert it into chemical energy

photosystem I

integral pigment and protein complex in thylakoid membranes that uses light energy to transport electrons from plastocyanin to NADP^+ (which becomes reduced to NADPH in the process)

photosystem II

integral protein and pigment complex in thylakoid membranes that transports electrons from water to the electron transport chain; oxygen is a product of PSII

primary electron acceptor

pigment or other organic molecule in the reaction center that accepts an energized electron from the reaction center

reaction center

complex of chlorophyll molecules and other organic molecules that is assembled around a special pair of chlorophyll molecules and a primary electron acceptor; capable of undergoing oxidation and reduction

spectrophotometer

instrument that can measure transmitted light and compute the absorption

wavelength

distance between consecutive points of equal position (two crests or two troughs) of a wave in a graphic representation; inversely proportional to the energy of the radiation

Using Light Energy to Make Organic Molecules

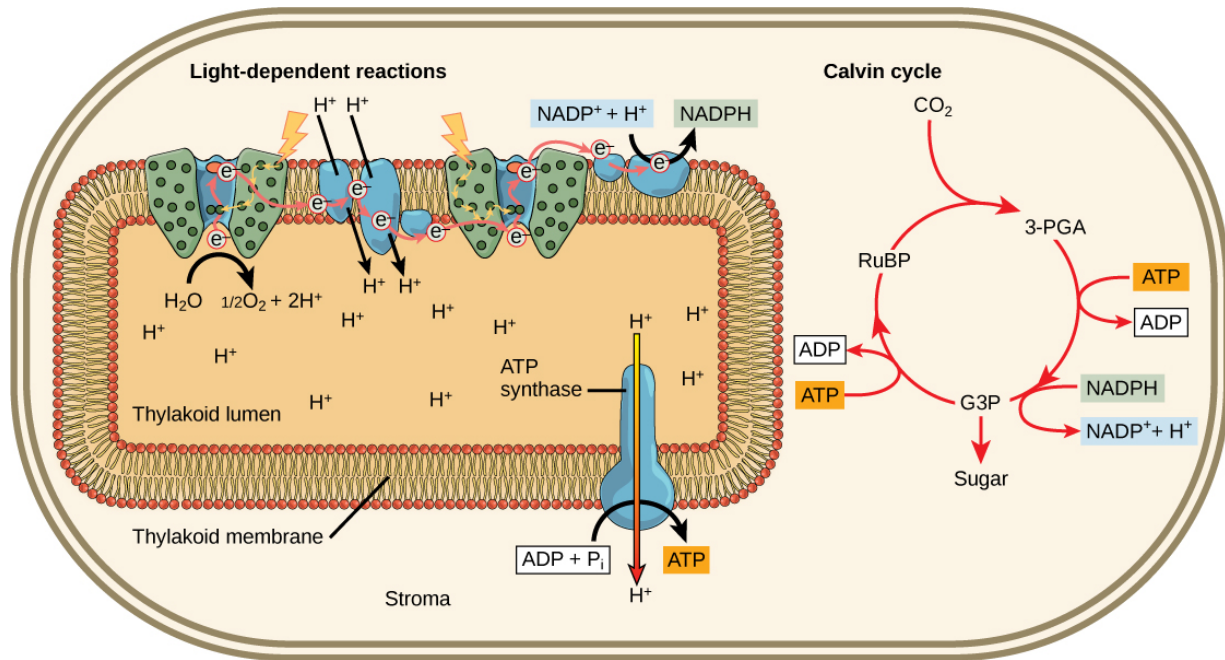
By the end of this section, you will be able to do the following:

- Describe the Calvin cycle
- Define carbon fixation
- Explain how photosynthesis works in the energy cycle of all living organisms

After the energy from the sun is converted into chemical energy and temporarily stored in ATP and NADPH molecules, the cell has the fuel needed to build carbohydrate molecules for long-term energy storage. The products of the light-dependent reactions, ATP and NADPH, have lifespans in the range of millionths of seconds, whereas the products of the light-independent reactions (carbohydrates and other forms of reduced carbon) can survive almost indefinitely. The carbohydrate molecules made will have a backbone of carbon atoms. But where does the carbon come from? It comes from carbon dioxide—the gas that is a waste product of respiration in microbes, fungi, plants, and animals.

The Calvin Cycle

In plants, carbon dioxide (CO₂) enters the leaves through stomata, where it diffuses over short distances through intercellular spaces until it reaches the mesophyll cells. Once in the mesophyll cells, CO₂ diffuses into the stroma of the chloroplast—the site of light-independent reactions of photosynthesis. These reactions actually have several names associated with them. Another term, the **Calvin cycle**, is named for the man who discovered it, and because these reactions function as a cycle. Others call it the Calvin-Benson cycle to include the name of another scientist involved in its discovery. The most outdated name is “dark reaction,” because light is not directly required ([\[link\]](#)). However, the term dark reaction can be misleading because it implies incorrectly that the reaction only occurs at night or is independent of light, which is why most scientists and instructors no longer use it.



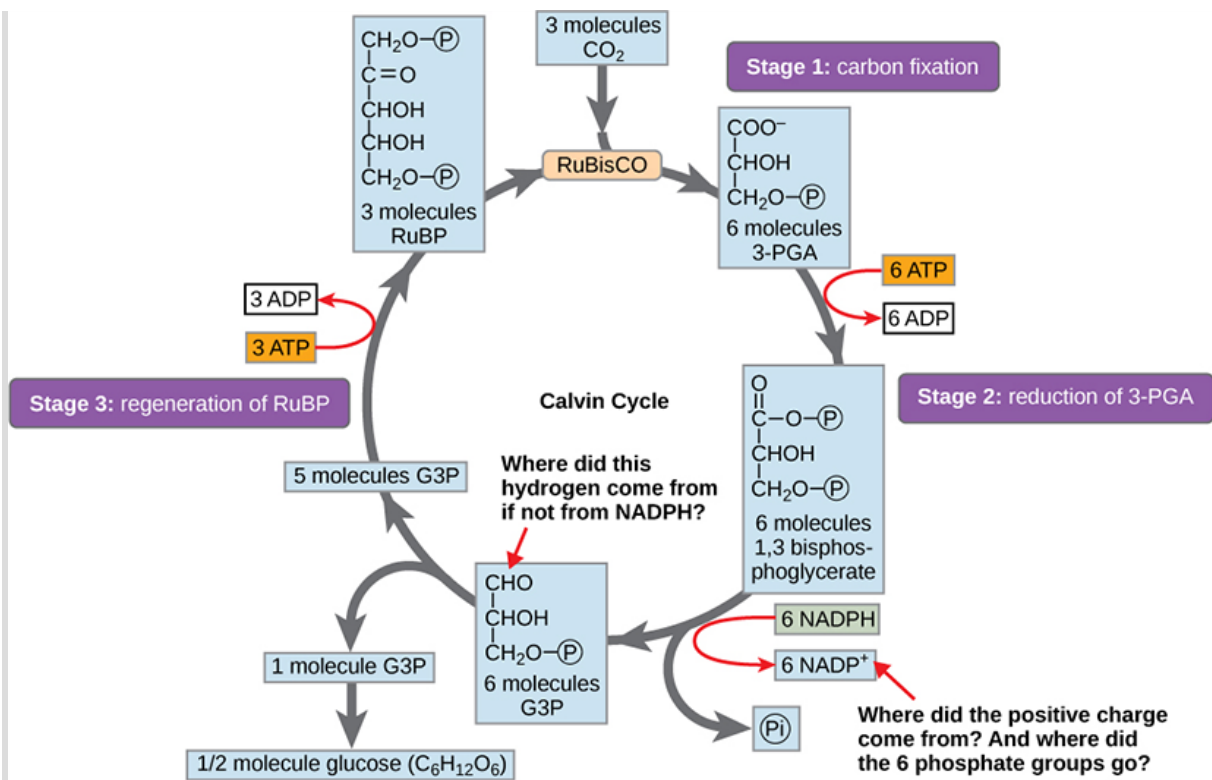
Light reactions harness energy from the sun to produce chemical bonds, ATP, and NADPH. These energy-carrying molecules are made in the stroma where carbon fixation takes place.

The light-independent reactions of the Calvin cycle can be organized into three basic stages: *fixation*, *reduction*, and *regeneration*.

Stage 1: Fixation

In the stroma, in addition to CO_2 , two other components are present to initiate the light-independent reactions: an enzyme called ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), and three molecules of ribulose bisphosphate (RuBP), as shown in [\[link\]](#). RuBP has five atoms of carbon, flanked by two phosphates.

Note:
Visual Connection



The Calvin cycle has three stages. In stage 1, the enzyme RuBisCO incorporates carbon dioxide into an organic molecule, 3-PGA. In stage 2, the organic molecule is reduced using electrons supplied by NADPH. In stage 3, RuBP, the molecule that starts the cycle, is regenerated so that the cycle can continue. Only one carbon dioxide molecule is incorporated at a time, so the cycle must be completed three times to produce a single three-carbon G3P molecule, and six times to produce a six-carbon glucose molecule.

Which of the following statements is true?

- In photosynthesis, oxygen, carbon dioxide, ATP, and NADPH are reactants. G3P and water are products.
- In photosynthesis, chlorophyll, water, and carbon dioxide are reactants. G3P and oxygen are products.
- In photosynthesis, water, carbon dioxide, ATP, and NADPH are reactants. RuBP and oxygen are products.

d. In photosynthesis, water and carbon dioxide are reactants. G3P and oxygen are products.

RuBisCO catalyzes a reaction between CO_2 and RuBP. For each CO_2 molecule that reacts with one RuBP, two molecules of another compound 3-phospho glyceric acid (3-PGA) form. PGA has three carbons and one phosphate. Each turn of the cycle involves only one RuBP and one carbon dioxide and forms two molecules of 3-PGA. The number of carbon atoms remains the same, as the atoms move to form new bonds during the reactions (3 C atoms from 3CO_2 + 15 C atoms from 3RuBP = 18 C atoms in 6 molecules of 3-PGA). This process is called **carbon fixation**, because CO_2 is “fixed” from an inorganic form into organic molecules.

Stage 2: Reduction

ATP and NADPH are used to convert the six molecules of 3-PGA into six molecules of a chemical called glyceraldehyde 3-phosphate (G3P). That is a reduction reaction because it involves the gain of electrons by 3-PGA. (Recall that a **reduction** is the gain of an electron by an atom or molecule.) Six molecules of both ATP and NADPH are used. For ATP, energy is released with the loss of the terminal phosphate atom, converting it into ADP; for NADPH, both energy and a hydrogen atom are lost, converting it into NADP^+ . Both of these molecules return to the nearby light-dependent reactions to be reused and re-energized.

Stage 3: Regeneration

Interestingly, at this point, only one of the G3P molecules leaves the Calvin cycle and is sent to the cytoplasm to contribute to the formation of other compounds needed by the plant. Because the G3P exported from the chloroplast has three carbon atoms, it takes three “turns” of the Calvin cycle to fix enough net carbon to export one G3P. But each turn makes two G3Ps,

thus three turns make six G3Ps. One is exported while the remaining five G3P molecules remain in the cycle and are used to regenerate RuBP, which enables the system to prepare for more CO₂ to be fixed. Three more molecules of ATP are used in these regeneration reactions.

Note:

Link to Learning

This [link](#) leads to an animation of photosynthesis and the Calvin cycle.

Note:

Evolution Connection

Photosynthesis

During the evolution of photosynthesis, a major shift occurred from the bacterial type of photosynthesis that involves only one photosystem and is typically anoxygenic (does not generate oxygen) into modern oxygenic (does generate oxygen) photosynthesis, employing two photosystems. This modern oxygenic photosynthesis is used by many organisms—from giant tropical leaves in the rainforest to tiny cyanobacterial cells—and the process and components of this photosynthesis remain largely the same. Photosystems absorb light and use electron transport chains to convert energy into the chemical energy of ATP and NADH. The subsequent light-independent reactions then assemble carbohydrate molecules with this energy.

In the harsh dry heat of the desert, plants must conserve every drop of water must be used to survive. Because stomata must open to allow for the uptake of CO₂, water escapes from the leaf during active photosynthesis. Desert plants have evolved processes to conserve water and deal with harsh conditions. Mechanisms to capture and store CO₂ allows plants to adapt to living with less water. Some plants such as cacti ([link](#)) can prepare materials for photosynthesis during the night by a temporary carbon fixation/storage process, because opening the stomata at this time conserves water due to cooler temperatures. During the day cacti use the captured CO₂ for photosynthesis, and keep their stomata closed.



The harsh conditions of the desert have led plants like these cacti to evolve variations of the light-independent reactions of photosynthesis. These variations increase the efficiency of water usage, helping to conserve water and energy. (credit: Piotr Wojtkowski)

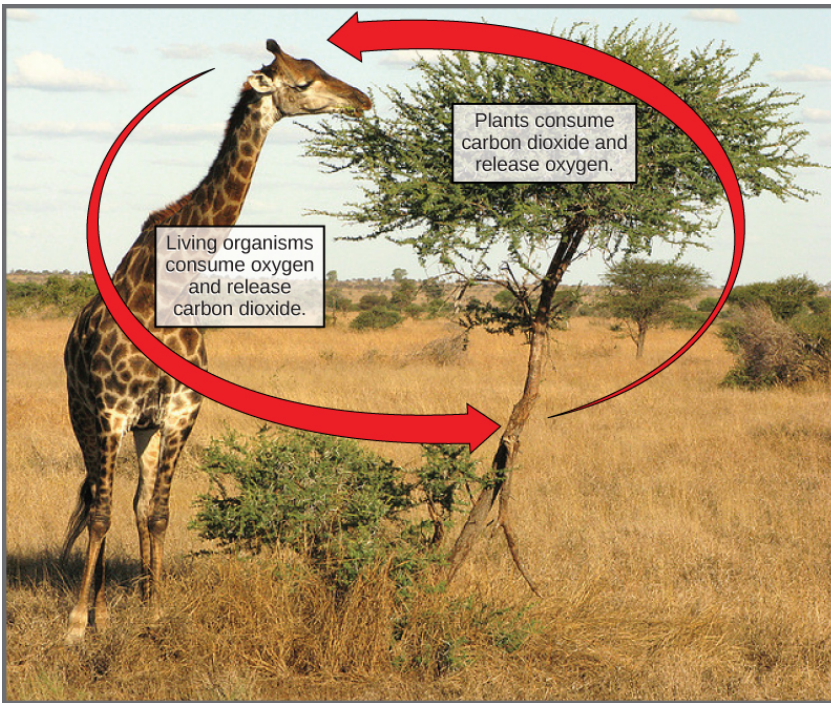
The Energy Cycle

Whether the organism is a bacterium, plant, or animal, all living things access energy by breaking down carbohydrate and other carbon-rich organic molecules. But if plants make carbohydrate molecules, why would they need to break them down, especially when it has been shown that the gas organisms release as a “waste product” (CO_2) acts as a substrate for the

formation of more food in photosynthesis? Remember, living things need energy to perform life functions. In addition, an organism can either make its own food or eat another organism—either way, the food still needs to be broken down. Finally, in the process of breaking down food, called cellular respiration, heterotrophs release needed energy and produce “waste” in the form of CO₂ gas.

However, in nature, there is no such thing as “waste.” Every single atom of matter and energy is conserved, recycled over and over infinitely. Substances change form or move from one type of molecule to another, but their constituent atoms never disappear ([\[link\]](#)).

In reality, CO₂ is no more a form of waste than oxygen is wasteful to photosynthesis. Both are byproducts of reactions that move on to other reactions. Photosynthesis absorbs light energy to build carbohydrates in chloroplasts, and aerobic cellular respiration releases energy by using oxygen to metabolize carbohydrates in the cytoplasm and mitochondria. Both processes use electron transport chains to capture the energy necessary to drive other reactions. These two powerhouse processes, photosynthesis and cellular respiration, function in biological, cyclical harmony to allow organisms to access life-sustaining energy that originates millions of miles away in a burning star humans call the sun.



Photosynthesis consumes carbon dioxide and produces oxygen. Aerobic respiration consumes oxygen and produces carbon dioxide. These two processes play an important role in the carbon cycle. (credit: modification of work by Stuart Bassil)

Section Summary

Using the energy carriers formed in the first steps of photosynthesis, the light-independent reactions, or the Calvin cycle, take in CO_2 from the atmosphere. An enzyme, RuBisCO, catalyzes a reaction with CO_2 and another organic compound, RuBP. After three cycles, a three-carbon molecule of G3P leaves the cycle to become part of a carbohydrate molecule. The remaining G3P molecules stay in the cycle to be regenerated into RuBP, which is then ready to react with more CO_2 . Photosynthesis forms an energy cycle with the process of cellular respiration. Because plants contain both chloroplasts and mitochondria, they rely upon both

photosynthesis and respiration for their ability to function in both the light and dark, and to be able to interconvert essential metabolites.

Visual Connection Questions

Exercise:

Problem: [\[link\]](#) Which of the following statements is true?

- a. In photosynthesis, oxygen, carbon dioxide, ATP, and NADPH are reactants. G3P and water are products.
- b. In photosynthesis, chlorophyll, water, and carbon dioxide are reactants. G3P and oxygen are products.
- c. In photosynthesis, water, carbon dioxide, ATP, and NADPH are reactants. RuBP and oxygen are products.
- d. In photosynthesis, water and carbon dioxide are reactants. G3P and oxygen are products.

Solution:

[\[link\]](#) D

Review Questions

Exercise:

Problem:

Which molecule must enter the Calvin cycle continually for the light-independent reactions to take place?

- a. RuBisCO
 - b. RuBP
 - c. 3-PGA
 - d. CO₂
-

Solution:

D

Exercise:

Problem:

Which order of molecular conversions is correct for the Calvin cycle?

- a. $\text{RuBP} + \text{G3P} \rightarrow 3\text{-PGA} \rightarrow \text{sugar}$
- b. $\text{RuBisCO} \rightarrow \text{CO}_2 \rightarrow \text{RuBP} \rightarrow \text{G3P}$
- c. $\text{RuBP} + \text{CO}_2 \rightarrow [\text{RuBisCO}] \text{ 3-PGA} \rightarrow \text{G3P}$
- d. $\text{CO}_2 \rightarrow 3\text{-PGA} \rightarrow \text{RuBP} \rightarrow \text{G3P}$

Solution:

C

Exercise:

Problem: Where in eukaryotic cells does the Calvin cycle take place?

- a. thylakoid membrane
- b. thylakoid lumen
- c. chloroplast stroma
- d. granum

Solution:

C

Exercise:

Problem: Which statement correctly describes carbon fixation?

- a. the conversion of CO_2 into an organic compound

- b. the use of RuBisCO to form 3-PGA
 - c. the production of carbohydrate molecules from G3P
 - d. the formation of RuBP from G3P molecules
 - e. the use of ATP and NADPH to reduce CO₂
-

Solution:

A

Exercise:

Problem:

If four molecules of carbon dioxide enter the Calvin cycle (four “turns” of the cycle), how many G3P molecules are produced and how many are exported?

- a. 4 G3P made, 1 G3P exported
 - b. 4 G3P made, 2 G3P exported
 - c. 8 G3P made, 1 G3P exported
 - d. 8 G3P made, 4 G3P exported
-

Solution:

C

Critical Thinking Questions

Exercise:

Problem:

Why is the third stage of the Calvin cycle called the regeneration stage?

Solution:

Because RuBP, the molecule needed at the start of the cycle, is regenerated from G3P.

Exercise:

Problem:

Which part of the light-independent reactions would be affected if a cell could not produce the enzyme RuBisCO?

Solution:

None of the cycle could take place, because RuBisCO is essential in fixing carbon dioxide. Specifically, RuBisCO catalyzes the reaction between carbon dioxide and RuBP at the start of the cycle.

Exercise:

Problem:

Why does it take three turns of the Calvin cycle to produce G3P, the initial product of photosynthesis?

Solution:

Because G3P has three carbon atoms, and each turn of the cycle takes in one carbon atom in the form of carbon dioxide.

Exercise:

Problem:

Imagine a sealed terrarium containing a plant and a beetle. How does each organism provide resources for the other? Could each organism survive if it was the only living thing in the terrarium? Why or why not?

Solution:

An energy cycle between a plant and a beetle would be as follows:

- A. Plant consumes carbon dioxide and releases oxygen as a byproduct of photosynthesis
- B. Beetle consumes oxygen and releases carbon dioxide to create chemical energy during aerobic respiration
- C. Plant takes up carbon dioxide from the air
- D. Repeat cycle

The plant would also provide a carbon-based food source for the beetle.

1. The beetle is a heterotroph, and would not survive without the plant because it would deplete all the oxygen within the terrarium.
2. The plant is an autotroph and could survive without the beetle, but it would be unlikely to grow. Through photosynthesis, the plant can make and store its own energy in carbon-based molecules, and produce oxygen. The oxygen can then be used to power aerobic respiration in the plant, which releases carbon dioxide. However, since the plant essentially continues to reuse its own resources cycling between carbon- and oxygen-consuming pathways, its growth would be limited.

Exercise:

Problem:

Compare the flow of energy with the flow of nutrients in a closed, sunny ecosystem consisting of a giraffe and a tree.

Solution:

In the defined ecosystem, energy would radiate from the Sun, and be absorbed by the chlorophyll in the leaves of the tree. Photosynthesis would occur in the leaves, transforming the light energy into stored chemical energy in the covalent bonds of carbon molecules. The giraffe would eat the leaves of the tree, and digest the carbon molecules to release energy.

In the same ecosystem, nutrients would cycle between the tree and the giraffe. The giraffe would consume oxygen and release carbon dioxide as its cells perform aerobic respiration to create chemical energy. The tree will consume the released carbon dioxide during photosynthesis to create its own stored chemical energy, and release oxygen as a by-product.

Glossary

Calvin cycle

light-independent reactions of photosynthesis that convert carbon dioxide from the atmosphere into carbohydrates using the energy and reducing power of ATP and NADPH

carbon fixation

process of converting inorganic CO₂ gas into organic compounds

reduction

gain of electron(s) by an atom or molecule

Nutrition and Energy Production

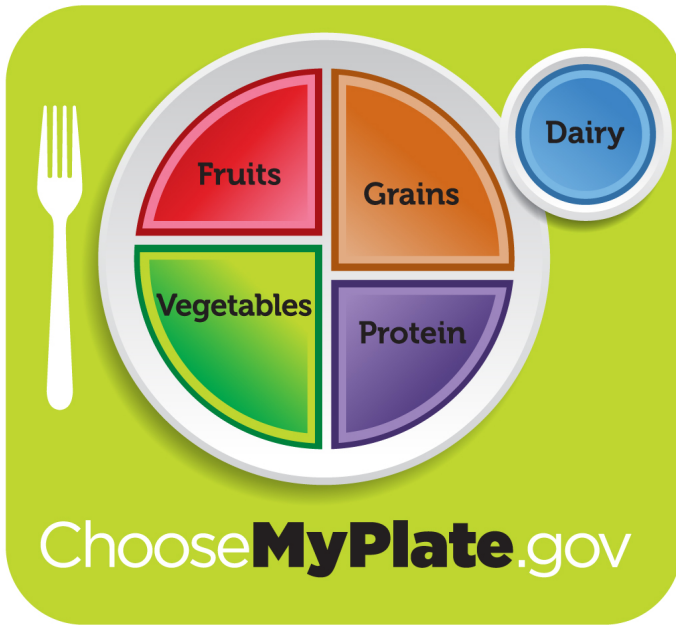
By the end of this section, you will be able to do the following:

- Explain why an animal's diet should be balanced and meet the needs of the body
- Define the primary components of food
- Describe the essential nutrients required for cellular function that cannot be synthesized by the animal body
- Explain how energy is produced through diet and digestion
- Describe how excess carbohydrates and energy are stored in the body

Given the diversity of animal life on our planet, it is not surprising that the animal diet would also vary substantially. The animal diet is the source of materials needed for building DNA and other complex molecules needed for growth, maintenance, and reproduction; collectively these processes are called biosynthesis. The diet is also the source of materials for ATP production in the cells. The diet must be balanced to provide the minerals and vitamins that are required for cellular function.

Food Requirements

What are the fundamental requirements of the animal diet? The animal diet should be well balanced and provide nutrients required for bodily function and the minerals and vitamins required for maintaining structure and regulation necessary for good health and reproductive capability. These requirements for a human are illustrated graphically in [\[link\]](#)



For humans, a balanced diet includes fruits, vegetables, grains, and protein.
(credit: USDA)

Note:

Link to Learning

The first step in ensuring that you are meeting the food requirements of your body is an awareness of the food groups and the nutrients they provide. To learn more about each food group and the recommended daily amounts, explore this [interactive site](#) by the United States Department of Agriculture.

Note:

Everyday Connection

Let's Move! Campaign

Obesity is a growing epidemic and the rate of obesity among children is rapidly rising in the United States. To combat childhood obesity and ensure that children get a healthy start in life, first lady Michelle Obama has

launched the Let's Move! campaign. The goal of this campaign is to educate parents and caregivers on providing healthy nutrition and encouraging active lifestyles to future generations. This program aims to involve the entire community, including parents, teachers, and healthcare providers to ensure that children have access to healthy foods—more fruits, vegetables, and whole grains—and consume fewer calories from processed foods. Another goal is to ensure that children get physical activity. With the increase in television viewing and stationary pursuits such as video games, sedentary lifestyles have become the norm. Learn more at <https://letsmove.obamawhitehouse.archives.gov>.

Organic Precursors

The organic molecules required for building cellular material and tissues must come from food. Carbohydrates or sugars are the primary source of organic carbons in the animal body. During digestion, digestible carbohydrates are ultimately broken down into glucose and used to provide energy through metabolic pathways. Complex carbohydrates, including polysaccharides, can be broken down into glucose through biochemical modification; however, humans do not produce the enzyme cellulase and lack the ability to derive glucose from the polysaccharide cellulose. In humans, these molecules provide the fiber required for moving waste through the large intestine and a healthy colon. The intestinal flora in the human gut are able to extract some nutrition from these plant fibers. The excess sugars in the body are converted into glycogen and stored in the liver and muscles for later use. Glycogen stores are used to fuel prolonged exertions, such as long-distance running, and to provide energy during food shortage. Excess glycogen can be converted to fats, which are stored in the lower layer of the skin of mammals for insulation and energy storage. Excess digestible carbohydrates are stored by mammals in order to survive famine and aid in mobility.

Another important requirement is that of nitrogen. Protein catabolism provides a source of organic nitrogen. Amino acids are the building blocks of proteins and protein breakdown provides amino acids that are used for cellular function. The carbon and nitrogen derived from these become the building block for nucleotides, nucleic acids, proteins, cells, and tissues.

Excess nitrogen must be excreted as it is toxic. Fats add flavor to food and promote a sense of satiety or fullness. Fatty foods are also significant sources of energy because one gram of fat contains nine calories. Fats are required in the diet to aid the absorption of fat-soluble vitamins and the production of fat-soluble hormones.

Essential Nutrients

While the animal body can synthesize many of the molecules required for function from the organic precursors, there are some nutrients that need to be consumed from food. These nutrients are termed **essential nutrients**, meaning they must be eaten, and the body cannot produce them.

The omega-3 alpha-linolenic acid and the omega-6 linoleic acid are essential fatty acids needed to make some membrane phospholipids. **Vitamins** are another class of essential organic molecules that are required in small quantities for many enzymes to function and, for this reason, are considered to be coenzymes. Absence or low levels of vitamins can have a dramatic effect on health, as outlined in [\[link\]](#) and [\[link\]](#). Both fat-soluble and water-soluble vitamins must be obtained from food. **Minerals**, listed in [\[link\]](#), are inorganic essential nutrients that must be obtained from food. Among their many functions, minerals help in structure and regulation and are considered cofactors. Certain amino acids also must be procured from food and cannot be synthesized by the body. These amino acids are the “essential” amino acids. The human body can synthesize only 11 of the 20 required amino acids; the rest must be obtained from food. The essential amino acids are listed in [\[link\]](#).

Water-soluble Essential Vitamins

Water-soluble Vitamin	Essential Vitamins Function	Deficiencies Can Lead To	Sources
------------------------------	------------------------------------	---------------------------------	----------------

Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin B ₁ (Thiamine)	Needed by the body to process lipids, proteins, and carbohydrates; coenzyme removes CO ₂ from organic compounds	Muscle weakness, Beriberi: reduced heart function, CNS problems	Milk, meat, dried beans, whole grains
Vitamin B ₂ (Riboflavin)	Takes an active role in metabolism, aiding in the conversion of food to energy (FAD and FMN)	Cracks or sores on the outer surface of the lips (cheilosis); inflammation and redness of the tongue; moist, scaly skin inflammation (seborrheic dermatitis)	Meat, eggs, enriched grains, vegetables

Water-soluble Essential Vitamins

Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin B ₃ (Niacin)	Used by the body to release energy from carbohydrates and to process alcohol; required for the synthesis of sex hormones; component of coenzyme NAD ⁺ and NADP ⁺	Pellagra, which can result in dermatitis, diarrhea, dementia, and death	Meat, eggs, grains, nuts, potatoes
Vitamin B ₅ (Pantothenic acid)	Assists in producing energy from foods (lipids, in particular); component of coenzyme A	Fatigue, poor coordination, retarded growth, numbness, tingling of hands and feet	Meat, whole grains, milk, fruits, vegetables

Water-soluble Essential Vitamins

Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin B ₆ (Pyridoxine)	The principal vitamin for processing amino acids and lipids; also helps convert nutrients into energy	Irritability, depression, confusion, mouth sores or ulcers, anemia, muscular twitching	Meat, dairy products, whole grains, orange juice
Vitamin B ₇ (Biotin)	Used in energy and amino acid metabolism, fat synthesis, and fat breakdown; helps the body use blood sugar	Hair loss, dermatitis, depression, numbness and tingling in the extremities; neuromuscular disorders	Meat, eggs, legumes and other vegetables
Vitamin B ₉ (Folic acid)	Assists the normal development of cells, especially during fetal development; helps metabolize nucleic and amino acids	Deficiency during pregnancy is associated with birth defects, such as neural tube defects and anemia	Leafy green vegetables, whole wheat, fruits, nuts, legumes

Water-soluble Essential Vitamins

Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin B ₁₂ (Cobalamin)	Maintains healthy nervous system and assists with blood cell formation; coenzyme in nucleic acid metabolism	Anemia, neurological disorders, numbness, loss of balance	Meat, eggs, animal products
Vitamin C (Ascorbic acid)	Helps maintain connective tissue: bone, cartilage, and dentin; boosts the immune system	Scurvy, which results in bleeding, hair and tooth loss; joint pain and swelling; delayed wound healing	Citrus fruits, broccoli, tomatoes, red sweet bell peppers

Fat-soluble Essential Vitamins

Vitamin	Function	Deficiencies Can Lead To	Sources
---------	----------	--------------------------	---------

Fat-soluble Essential Vitamins			
Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin A (Retinol)	Critical to the development of bones, teeth, and skin; helps maintain eyesight, enhances the immune system, fetal development, gene expression	Night-blindness, skin disorders, impaired immunity	Dark green leafy vegetables, yellow-orange vegetables, fruits, milk, butter
Vitamin D	Critical for calcium absorption for bone development and strength; maintains a stable nervous system; maintains a normal and strong heartbeat; helps in blood clotting	Rickets, osteomalacia, immunity	Cod liver oil, milk, egg yolk

Fat-soluble Essential Vitamins			
Vitamin	Function	Deficiencies Can Lead To	Sources
Vitamin E (Tocopherol)	Lessens oxidative damage of cells and prevents lung damage from pollutants; vital to the immune system	Deficiency is rare; anemia, nervous system degeneration	Wheat germ oil, unrefined vegetable oils, nuts, seeds, grains
Vitamin K (Phylloquinone)	Essential to blood clotting	Bleeding and easy bruising	Leafy green vegetables, tea



A healthy diet should include a variety of

foods to ensure that needs for essential nutrients are met. (credit: Keith Weller, USDA ARS)

Minerals and Their Function in the Human Body			
Mineral	Function	Deficiencies Can Lead To	Sources
*Calcium	Needed for muscle and neuron function; heart health; builds bone and supports synthesis and function of blood cells; nerve function	Osteoporosis, rickets, muscle spasms, impaired growth	Milk, yogurt, fish, green leafy vegetables, legumes
*Chlorine	Needed for production of hydrochloric acid (HCl) in the stomach and nerve function; osmotic balance	Muscle cramps, mood disturbances, reduced appetite	Table salt

Minerals and Their Function in the Human Body			
Mineral	Function	Deficiencies Can Lead To	Sources
Copper (trace amounts)	Required component of many redox enzymes, including cytochrome c oxidase; cofactor for hemoglobin synthesis	Copper deficiency is rare	Liver, oysters, cocoa, chocolate, sesame, nuts
Iodine	Required for the synthesis of thyroid hormones	Goiter	Seafood, iodized salt, dairy products
Iron	Required for many proteins and enzymes, notably hemoglobin, to prevent anemia	Anemia, which causes poor concentration, fatigue, and poor immune function	Red meat, leafy green vegetables, fish (tuna, salmon), eggs, dried fruits, beans, whole grains

Minerals and Their Function in the Human Body			
Mineral	Function	Deficiencies Can Lead To	Sources
*Magnesium	Required cofactor for ATP formation; bone formation; normal membrane functions; muscle function	Mood disturbances, muscle spasms	Whole grains, leafy green vegetables
Manganese (trace amounts)	A cofactor in enzyme functions; trace amounts are required	Manganese deficiency is rare	Common in most foods
Molybdenum (trace amounts)	Acts as a cofactor for three essential enzymes in humans: sulfite oxidase, xanthine oxidase, and aldehyde oxidase	Molybdenum deficiency is rare	
*Phosphorus	A component of bones and teeth; helps regulate acid-base balance; nucleotide synthesis	Weakness, bone abnormalities, calcium loss	Milk, hard cheese, whole grains, meats

Minerals and Their Function in the Human Body			
Mineral	Function	Deficiencies Can Lead To	Sources
*Potassium	Vital for muscles, heart, and nerve function	Cardiac rhythm disturbance, muscle weakness	Legumes, potato skin, tomatoes, bananas
Selenium (trace amounts)	A cofactor essential to activity of antioxidant enzymes like glutathione peroxidase; trace amounts are required	Selenium deficiency is rare	Common in most foods
*Sodium	Systemic electrolyte required for many functions; acid-base balance; water balance; nerve function	Muscle cramps, fatigue, reduced appetite	Table salt
Zinc (trace amounts)	Required for several enzymes such as carboxypeptidase, liver alcohol dehydrogenase, and carbonic anhydrase	Anemia, poor wound healing, can lead to short stature	Common in most foods

Minerals and Their Function in the Human Body			
Mineral	Function	Deficiencies Can Lead To	Sources
*Greater than 200mg/day required			

Essential Amino Acids	
Amino acids that must be consumed	Amino acids anabolized by the body
isoleucine	alanine
leucine	selenocysteine
lysine	aspartate
methionine	cysteine
phenylalanine	glutamate
tryptophan	glycine
valine	proline
histidine*	serine
threonine	tyrosine
arginine*	asparagine

Essential Amino Acids	
Amino acids that must be consumed	Amino acids anabolized by the body
*The human body can synthesize histidine and arginine, but not in the quantities required, especially for growing children.	

Food Energy and ATP

Animals need food to obtain energy and maintain homeostasis. Homeostasis is the ability of a system to maintain a stable internal environment even in the face of external changes to the environment. For example, the normal body temperature of humans is 37°C (98.6°F). Humans maintain this temperature even when the external temperature is hot or cold. It takes energy to maintain this body temperature, and animals obtain this energy from food.

The primary source of energy for animals is carbohydrates, mainly glucose. Glucose is called the body's fuel. The digestible carbohydrates in an animal's diet are converted to glucose molecules through a series of catabolic chemical reactions.

Adenosine triphosphate, or ATP, is the primary energy currency in cells; ATP stores energy in phosphate ester bonds. ATP releases energy when the phosphodiester bonds are broken and ATP is converted to ADP and a phosphate group. ATP is produced by the oxidative reactions in the cytoplasm and mitochondrion of the cell, where carbohydrates, proteins, and fats undergo a series of metabolic reactions collectively called cellular respiration. For example, glycolysis is a series of reactions in which glucose is converted to pyruvic acid and some of its chemical potential energy is transferred to NADH and ATP.

ATP is required for all cellular functions. It is used to build the organic molecules that are required for cells and tissues; it provides energy for muscle contraction and for the transmission of electrical signals in the nervous system. When the amount of ATP is available in excess of the body's requirements, the liver uses the excess ATP and excess glucose to produce

molecules called glycogen. Glycogen is a polymeric form of glucose and is stored in the liver and skeletal muscle cells. When blood sugar drops, the liver releases glucose from stores of glycogen. Skeletal muscle converts glycogen to glucose during intense exercise. The process of converting glucose and excess ATP to glycogen and the storage of excess energy is an evolutionarily important step in helping animals deal with mobility, food shortages, and famine.

Note:

Everyday Connection

Obesity

Obesity is a major health concern in the United States, and there is a growing focus on reducing obesity and the diseases it may lead to, such as type-2 diabetes, cancers of the colon and breast, and cardiovascular disease. How does the food consumed contribute to obesity?

Fatty foods are calorie-dense, meaning that they have more calories per unit mass than carbohydrates or proteins. One gram of carbohydrates has four calories, one gram of protein has four calories, and one gram of fat has nine calories. Animals tend to seek lipid-rich food for their higher energy content. The signals of hunger (“time to eat”) and satiety (“time to stop eating”) are controlled in the hypothalamus region of the brain. Foods that are rich in fatty acids tend to promote satiety more than foods that are rich only in carbohydrates.

Excess carbohydrate and ATP are used by the liver to synthesize glycogen. The pyruvate produced during glycolysis is used to synthesize fatty acids. When there is more glucose in the body than required, the resulting excess pyruvate is converted into molecules that eventually result in the synthesis of fatty acids within the body. These fatty acids are stored in adipose cells—the fat cells in the mammalian body whose primary role is to store fat for later use.

It is important to note that some animals benefit from obesity. Polar bears and seals need body fat for insulation and to keep them from losing body heat during Arctic winters. When food is scarce, stored body fat provides energy for maintaining homeostasis. Fats prevent famine in mammals, allowing them to access energy when food is not available on a daily basis; fats are stored when a large kill is made or lots of food is available.

Section Summary

Animal diet should be balanced and meet the needs of the body.

Carbohydrates, proteins, and fats are the primary components of food. Some essential nutrients are required for cellular function but cannot be produced by the animal body. These include vitamins, minerals, some fatty acids, and some amino acids. Food intake in more than necessary amounts is stored as glycogen in the liver and muscle cells, and in fat cells. Excess adipose storage can lead to obesity and serious health problems. ATP is the energy currency of the cell and is obtained from the metabolic pathways. Excess carbohydrates and energy are stored as glycogen in the body.

Review Questions

Exercise:

Problem: Which of the following statements is not true?

- a. Essential nutrients can be synthesized by the body.
- b. Vitamins are required in small quantities for bodily function.
- c. Some amino acids can be synthesized by the body, while others need to be obtained from diet.
- d. Vitamins come in two categories: fat-soluble and water-soluble.

Solution:

A

Exercise:

Problem: Which of the following is a water-soluble vitamin?

- a. vitamin A
 - b. vitamin E
 - c. vitamin K
 - d. vitamin C
-

Solution:

D

Exercise:

Problem: What is the primary fuel for the body?

- a. carbohydrates
- b. lipids
- c. protein
- d. glycogen

Solution:

A

Exercise:

Problem: Excess glucose is stored as _____.

- a. fat
- b. glucagon
- c. glycogen
- d. it is not stored in the body

Solution:

C

Exercise:

Problem:

Many distance runners “carb load” the day before a big race. How does this eating strategy provide an advantage to the runner?

- a. The carbohydrates cause the release of insulin.

- b. The excess carbohydrates are converted to fats, which have a higher calorie density.
 - c. The glucose from the carbohydrates lets the muscles make excess ATP overnight.
 - d. The excess carbohydrates can be stored in the muscles as glycogen.
-

Solution:

D

Critical Thinking Questions

Exercise:

Problem: What are essential nutrients?

Solution:

Essential nutrients are those nutrients that must be obtained from the diet because they cannot be produced by the body. Vitamins and minerals are examples of essential nutrients.

Exercise:

Problem: What is the role of minerals in maintaining good health?

Solution:

Minerals—such as potassium, sodium, and calcium—are required for the functioning of many cellular processes, including muscle contraction and nerve conduction. While minerals are required in trace amounts, not having minerals in the diet can be potentially harmful.

Exercise:

Problem: Discuss why obesity is a growing epidemic.

Solution:

In the United States, obesity, particularly childhood obesity, is a growing concern. Some of the contributors to this situation include sedentary lifestyles and consuming more processed foods and less fruits and vegetables. As a result, even young children who are obese can face health concerns.

Exercise:**Problem:**

There are several nations where malnourishment is a common occurrence. What may be some of the health challenges posed by malnutrition?

Solution:

Malnutrition, often in the form of not getting enough calories or not enough of the essential nutrients, can have severe consequences. Many malnourished children have vision and dental problems, and over the years may develop many serious health problems.

Exercise:**Problem:**

Generally describe how a piece of bread can power your legs as you walk up a flight of stairs.

Solution:

A piece of bread is eaten and converted into chemical energy. The bread is broken down in the mouth by mastication and salivary enzymes, then transferred to the stomach for further digestion. After digestion by the acid and digestive enzymes in the stomach, the macromolecules that made up the bread move into the small intestine. In the small intestine, the carbohydrates from the bread are absorbed through the microvilli into the bloodstream. In muscle cells in the legs, the carbohydrates can be broken down into glucose, and then used for cellular respiration to

create ATP. The muscles in the leg then use the ATP to perform the mechanical work needed to climb a flight of stairs.

Exercise:

Problem:

In the 1990s fat-free foods became popular among people trying to lose weight. However, many dieticians now conclude that the fat-free trend made people less healthy and heavier. Describe how this could occur.

Solution:

Fats are an essential component of a healthy diet, and needed by the body to function. Fats are essential for many processes, including the absorption of fat-soluble vitamins and production of some hormones. Fats also send a satiation signal to the brain that regulates hunger. Without fats in their diets many people may have actually consumed more calories, which would have resulted in weight gain.

Glossary

essential nutrient

nutrient that cannot be synthesized by the body; it must be obtained from food

mineral

inorganic, elemental molecule that carries out important roles in the body

vitamin

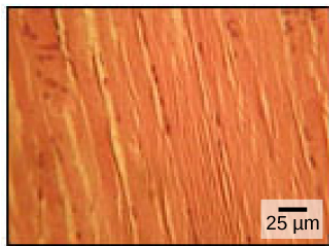
organic substance necessary in small amounts to sustain life

Muscle Contraction and Locomotion

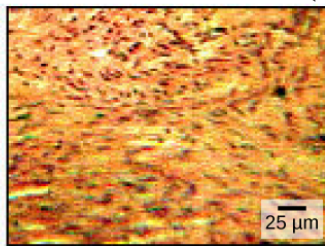
By the end of this section, you will be able to do the following:

- Classify the different types of muscle tissue
- Explain the role of muscles in locomotion

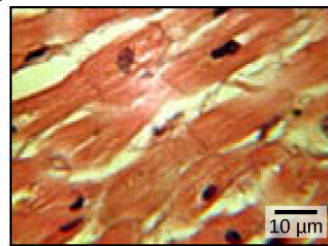
Muscle cells are specialized for contraction. Muscles allow for motions such as walking, and they also facilitate bodily processes such as respiration and digestion. The body contains three types of muscle tissue: skeletal muscle, cardiac muscle, and smooth muscle ([\[link\]](#)).



Skeletal muscle



Smooth muscle



Cardiac muscle

The body contains three types of muscle tissue: skeletal muscle, smooth muscle, and cardiac muscle, visualized here using light microscopy. Smooth muscle cells are short, tapered at each end, and have only one plump nucleus in each. Cardiac muscle cells are branched and striated, but short. The cytoplasm may branch, and they have one nucleus in the center of the cell. (credit: modification of work by NCI, NIH; scale-bar data from Matt Russell)

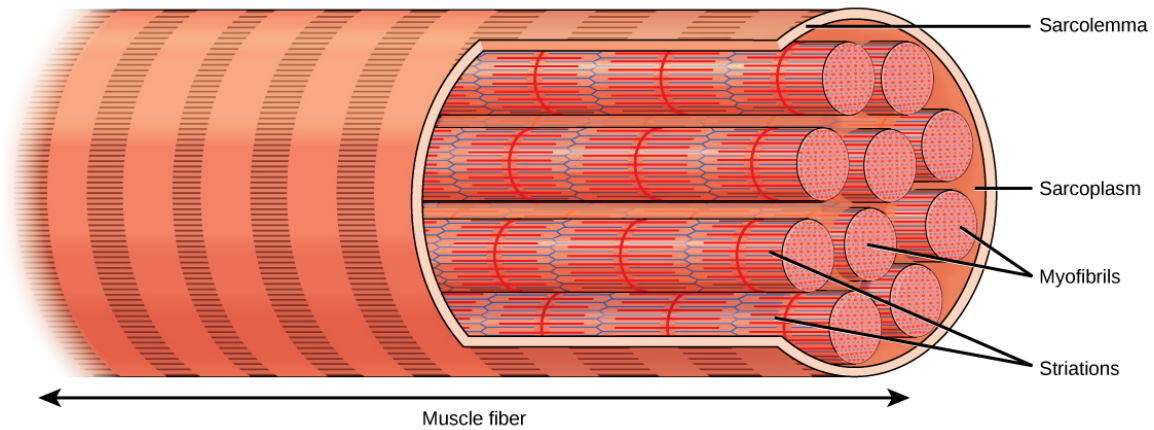
Skeletal muscle tissue forms skeletal muscles, which attach to bones or skin and control locomotion and any movement that can be consciously controlled. Because it can be controlled by thought, skeletal muscle is also called voluntary muscle. Skeletal muscles are long and cylindrical in appearance; when viewed under a microscope, skeletal muscle tissue has a striped or striated appearance. The striations are caused by the regular arrangement of contractile proteins (actin and myosin). **Actin** is a globular contractile protein that interacts with **myosin** for muscle contraction. Skeletal muscle also has multiple nuclei present in a single cell.

Smooth muscle tissue occurs in the walls of hollow organs such as the intestines, stomach, and urinary bladder, and around passages such as the respiratory tract and blood vessels. Smooth muscle has no striations, is not under voluntary control, has only one nucleus per cell, is tapered at both ends, and is called involuntary muscle.

Cardiac muscle tissue is only found in the heart, and cardiac contractions pump blood throughout the body and maintain blood pressure. Like skeletal muscle, cardiac muscle is striated, but unlike skeletal muscle, cardiac muscle cannot be consciously controlled and is called involuntary muscle. It has one nucleus per cell, is branched, and is distinguished by the presence of intercalated disks.

Skeletal Muscle Fiber Structure

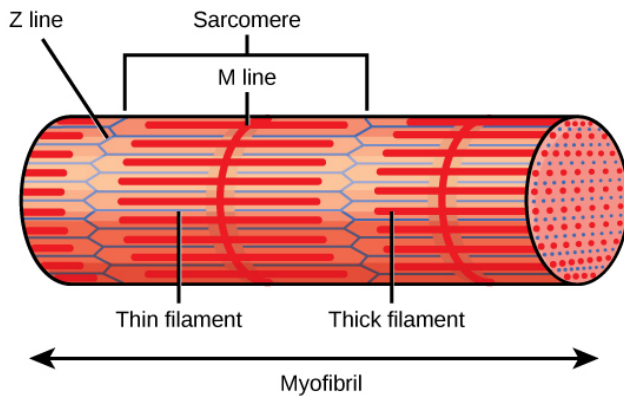
Each skeletal muscle fiber is a skeletal muscle cell. These cells are incredibly large, with diameters of up to 100 μm and lengths of up to 30 cm. The plasma membrane of a skeletal muscle fiber is called the **sarcolemma**. The sarcolemma is the site of action potential conduction, which triggers muscle contraction. Within each muscle fiber are **myofibrils**—long cylindrical structures that lie parallel to the muscle fiber. Myofibrils run the entire length of the muscle fiber, and because they are only approximately 1.2 μm in diameter, hundreds to thousands can be found inside one muscle fiber. They attach to the sarcolemma at their ends, so that as myofibrils shorten, the entire muscle cell contracts ([link](#)).



A skeletal muscle cell is surrounded by a plasma membrane called the sarcolemma with a cytoplasm called the sarcoplasm. A muscle fiber is composed of many fibrils, packaged into orderly units.

The striated appearance of skeletal muscle tissue is a result of repeating bands of the proteins actin and myosin that are present along the length of myofibrils. Dark A bands and light I bands repeat along myofibrils, and the alignment of myofibrils in the cell causes the entire cell to appear striated or banded.

Each I band has a dense line running vertically through the middle called a Z disc or Z line. The Z discs mark the border of units called **sarcomeres**, which are the functional units of skeletal muscle. One sarcomere is the space between two consecutive Z discs and contains one entire A band and two halves of an I band, one on either side of the A band. A myofibril is composed of many sarcomeres running along its length, and as the sarcomeres individually contract, the myofibrils and muscle cells shorten ([link](#)).



A sarcomere is the region from one Z line to the next Z line. Many sarcomeres are present in a myofibril, resulting in the striation pattern characteristic of skeletal muscle.

Myofibrils are composed of smaller structures called **myofilaments**. There are two main types of filaments: thick filaments and thin filaments; each has different compositions and locations. **Thick filaments** occur only in the A band of a myofibril. **Thin filaments** attach to a protein in the Z disc called alpha-actinin and occur across the entire length of the I band and partway into the A band. The region at which thick and thin filaments overlap has a dense appearance, as there is little space between the filaments. Thin filaments do not extend all the way into the A bands, leaving a central region of the A band that only contains thick filaments. This central region of the A band looks slightly lighter than the rest of the A band and is called the H zone. The middle of the H zone has a vertical line called the M line, at which accessory proteins hold together thick filaments. Both the Z disc and the M line hold myofilaments in place to maintain the structural arrangement and layering of the myofibril. Myofibrils are connected to each other by intermediate, or desmin, filaments that attach to the Z disc.

Thick and thin filaments are themselves composed of proteins. Thick filaments are composed of the protein myosin. The tail of a myosin molecule connects with other myosin molecules to form the central region of a thick filament near the M line, whereas the heads align on either side of the thick filament where the thin filaments overlap. The primary component of thin filaments is the actin protein. Two other components of the thin filament are tropomyosin and troponin. Actin has binding sites for myosin attachment. Strands of tropomyosin block the binding sites and prevent actin–myosin interactions when the muscles are at rest. Troponin consists of three globular subunits. One subunit binds to tropomyosin, one subunit binds to actin, and one subunit binds Ca^{2+} ions.

Note:

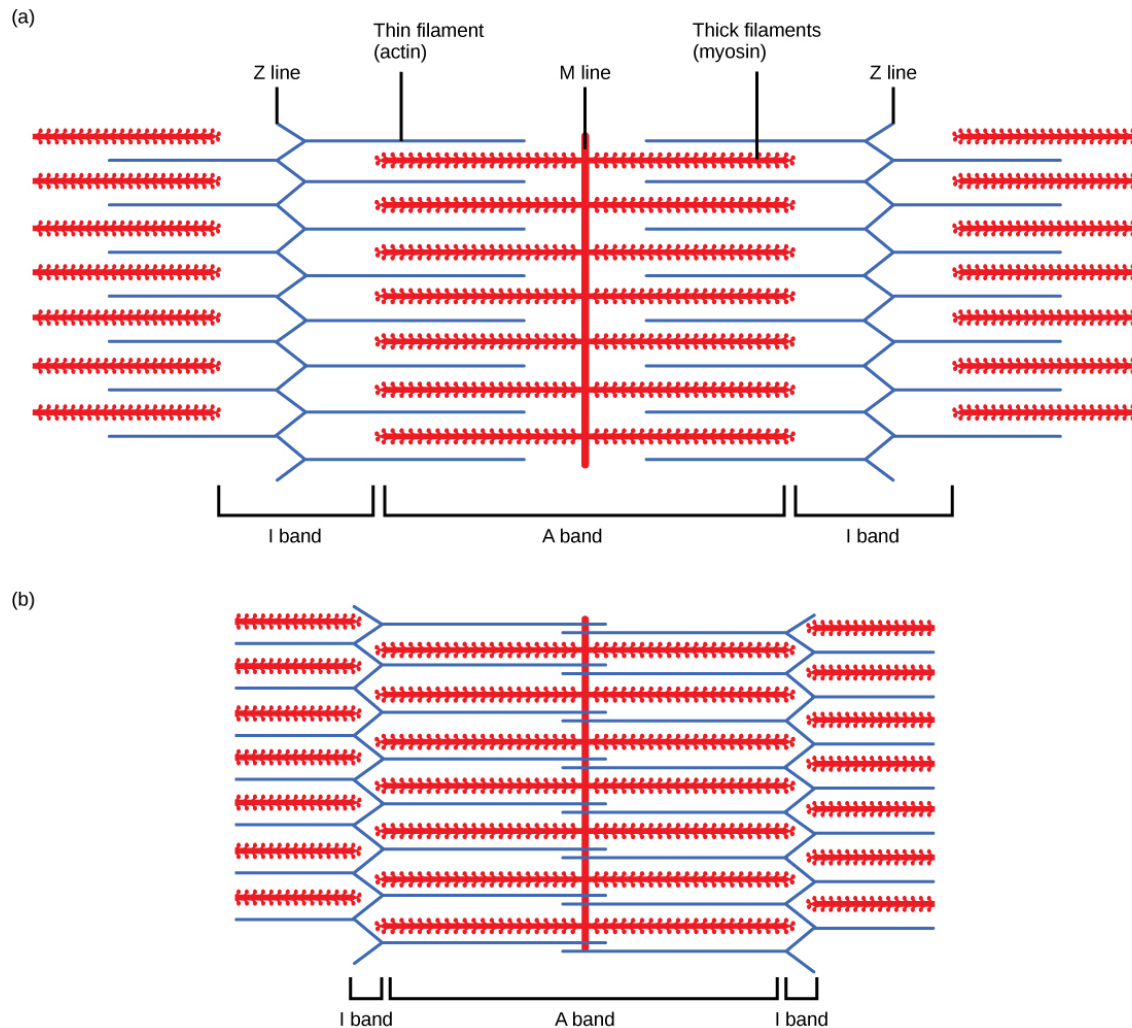
Link to Learning

View this animation showing the organization of muscle fibers.

https://www.openstax.org/l/skeletal_muscle

Sliding Filament Model of Contraction

For a muscle cell to contract, the sarcomere must shorten. However, thick and thin filaments—the components of sarcomeres—do not shorten. Instead, they slide by one another, causing the sarcomere to shorten while the filaments remain the same length. The sliding filament theory of muscle contraction was developed to fit the differences observed in the named bands on the sarcomere at different degrees of muscle contraction and relaxation. The mechanism of contraction is the binding of myosin to actin, forming cross-bridges that generate filament movement ([link](#)).



When (a) a sarcomere (b) contracts, the Z lines move closer together and the I band gets smaller. The A band stays the same width and, at full contraction, the thin filaments overlap.

When a sarcomere shortens, some regions shorten whereas others stay the same length. A sarcomere is defined as the distance between two consecutive Z discs or Z lines; when a muscle contracts, the distance between the Z discs is reduced. The H zone—the central region of the A zone—contains only thick filaments and is shortened during contraction. The I band contains only thin filaments and also shortens. The A band does not shorten—it remains the same length—but A bands of different sarcomeres move closer together during contraction, eventually

disappearing. Thin filaments are pulled by the thick filaments toward the center of the sarcomere until the Z discs approach the thick filaments. The zone of overlap, in which thin filaments and thick filaments occupy the same area, increases as the thin filaments move inward.

ATP and Muscle Contraction

The motion of muscle shortening occurs as myosin heads bind to actin and pull the actin inwards. This action requires energy, which is provided by ATP. Myosin binds to actin at a binding site on the globular actin protein. Myosin has another binding site for ATP at which enzymatic activity hydrolyzes ATP to ADP, releasing an inorganic phosphate molecule and energy.

ATP binding causes myosin to release actin, allowing actin and myosin to detach from each other. After this happens, the newly bound ATP is converted to ADP and inorganic phosphate, P_i . The enzyme at the binding site on myosin is called ATPase. The energy released during ATP hydrolysis changes the angle of the myosin head into a “cocked” position. The myosin head is then in a position for further movement, possessing potential energy, but ADP and P_i are still attached. If actin binding sites are covered and unavailable, the myosin will remain in the high energy configuration with ATP hydrolyzed, but still attached.

If the actin binding sites are uncovered, a cross-bridge will form; that is, the myosin head spans the distance between the actin and myosin molecules. P_i is then released, allowing myosin to expend the stored energy as a conformational change. The myosin head moves toward the M line, pulling the actin along with it. As the actin is pulled, the filaments move approximately 10 nm toward the M line. This movement is called the power stroke, as it is the step at which force is produced. As the actin is pulled toward the M line, the sarcomere shortens and the muscle contracts.

When the myosin head is “cocked,” it contains energy and is in a high-energy configuration. This energy is expended as the myosin head moves through the power stroke; at the end of the power stroke, the myosin head is in a low-energy position. After the power stroke, ADP is released; however,

the cross-bridge formed is still in place, and actin and myosin are bound together. ATP can then attach to myosin, which allows the cross-bridge cycle to start again and further muscle contraction can occur ([link]).

Note:

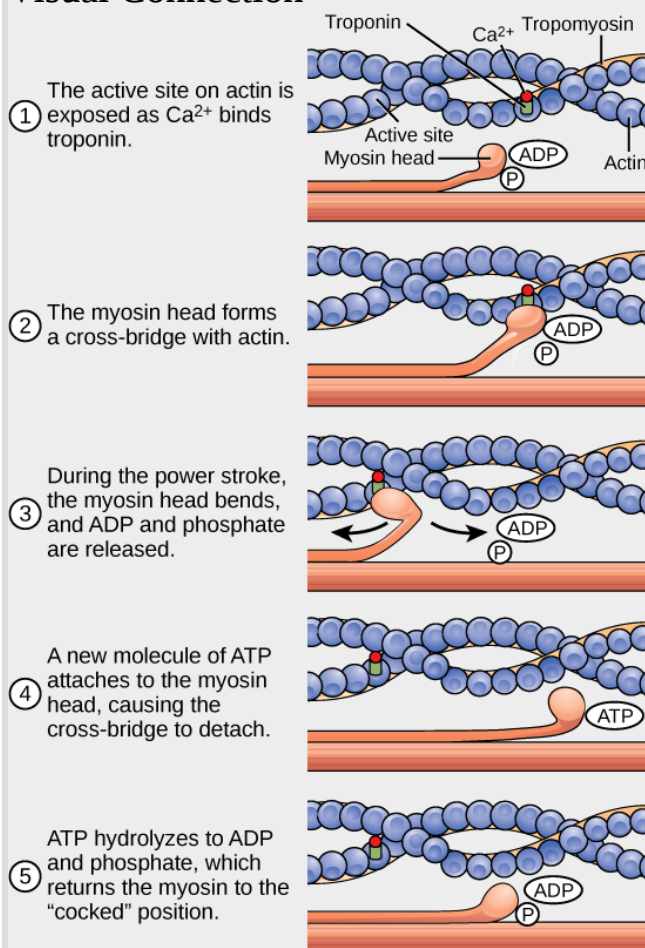
Link to Learning

Watch this video explaining how a muscle contraction is signaled.

https://www.openstax.org/l/contract_muscle

Note:

Visual Connection



The cross-bridge muscle

contraction cycle, which is triggered by Ca^{2+} binding to the actin active site, is shown. With each contraction cycle, actin moves relative to myosin.

Which of the following statements about muscle contraction is true?

- a. The power stroke occurs when ATP is hydrolyzed to ADP and phosphate.
- b. The power stroke occurs when ADP and phosphate dissociate from the myosin head.
- c. The power stroke occurs when ADP and phosphate dissociate from the actin active site.
- d. The power stroke occurs when Ca^{2+} binds the calcium head.

Note:

Link to Learning

View this [animation](#) of the cross-bridge muscle contraction.

Regulatory Proteins

When a muscle is in a resting state, actin and myosin are separated. To keep actin from binding to the active site on myosin, regulatory proteins block the molecular binding sites. **Tropomyosin** blocks myosin binding sites on actin molecules, preventing cross-bridge formation and preventing contraction in a muscle without nervous input. **Troponin** binds to tropomyosin and helps to position it on the actin molecule; it also binds calcium ions.

To enable a muscle contraction, tropomyosin must change conformation, uncovering the myosin-binding site on an actin molecule and allowing

cross-bridge formation. This can only happen in the presence of calcium, which is kept at extremely low concentrations in the sarcoplasm. If present, calcium ions bind to troponin, causing conformational changes in troponin that allow tropomyosin to move away from the myosin binding sites on actin. Once the tropomyosin is removed, a cross-bridge can form between actin and myosin, triggering contraction. Cross-bridge cycling continues until Ca^{2+} ions and ATP are no longer available and tropomyosin again covers the binding sites on actin.

Excitation–Contraction Coupling

Excitation–contraction coupling is the link (transduction) between the action potential generated in the sarcolemma and the start of a muscle contraction. The trigger for calcium release from the sarcoplasmic reticulum into the sarcoplasm is a neural signal. Each skeletal muscle fiber is controlled by a motor neuron, which conducts signals from the brain or spinal cord to the muscle. The area of the sarcolemma on the muscle fiber that interacts with the neuron is called the **motor end plate**. The end of the neuron's axon is called the synaptic terminal, and it does not actually contact the motor end plate. A small space called the synaptic cleft separates the synaptic terminal from the motor end plate. Electrical signals travel along the neuron's axon, which branches through the muscle and connects to individual muscle fibers at a neuromuscular junction.

The ability of cells to communicate electrically requires that the cells expend energy to create an electrical gradient across their cell membranes. This charge gradient is carried by ions, which are differentially distributed across the membrane. Each ion exerts an electrical influence and a concentration influence. Just as milk will eventually mix with coffee without the need to stir, ions also distribute themselves evenly, if they are permitted to do so. In this case, they are not permitted to return to an evenly mixed state.

The sodium–potassium ATPase uses cellular energy to move K^+ ions inside the cell and Na^+ ions outside. This alone accumulates a small electrical charge, but a big concentration gradient. There is lots of K^+ in the cell and lots of Na^+ outside the cell. Potassium is able to leave the cell through K^+

channels that are open 90% of the time, and it does. However, Na^+ channels are rarely open, so Na^+ remains outside the cell. When K^+ leaves the cell, obeying its concentration gradient, that effectively leaves a negative charge behind. So at rest, there is a large concentration gradient for Na^+ to enter the cell, and there is an accumulation of negative charges left behind in the cell. This is the resting membrane potential. Potential in this context means a separation of electrical charge that is capable of doing work. It is measured in volts, just like a battery. However, the transmembrane potential is considerably smaller (0.07 V); therefore, the small value is expressed as millivolts (mV) or 70 mV. Because the inside of a cell is negative compared with the outside, a minus sign signifies the excess of negative charges inside the cell, -70 mV .

If an event changes the permeability of the membrane to Na^+ ions, they will enter the cell. That will change the voltage. This is an electrical event, called an action potential, that can be used as a cellular signal.

Communication occurs between nerves and muscles through neurotransmitters. Neuron action potentials cause the release of neurotransmitters from the synaptic terminal into the synaptic cleft, where they can then diffuse across the synaptic cleft and bind to a receptor molecule on the motor end plate. The motor end plate possesses junctional folds—folds in the sarcolemma that create a large surface area for the neurotransmitter to bind to receptors. The receptors are actually sodium channels that open to allow the passage of Na^+ into the cell when they receive a neurotransmitter signal.

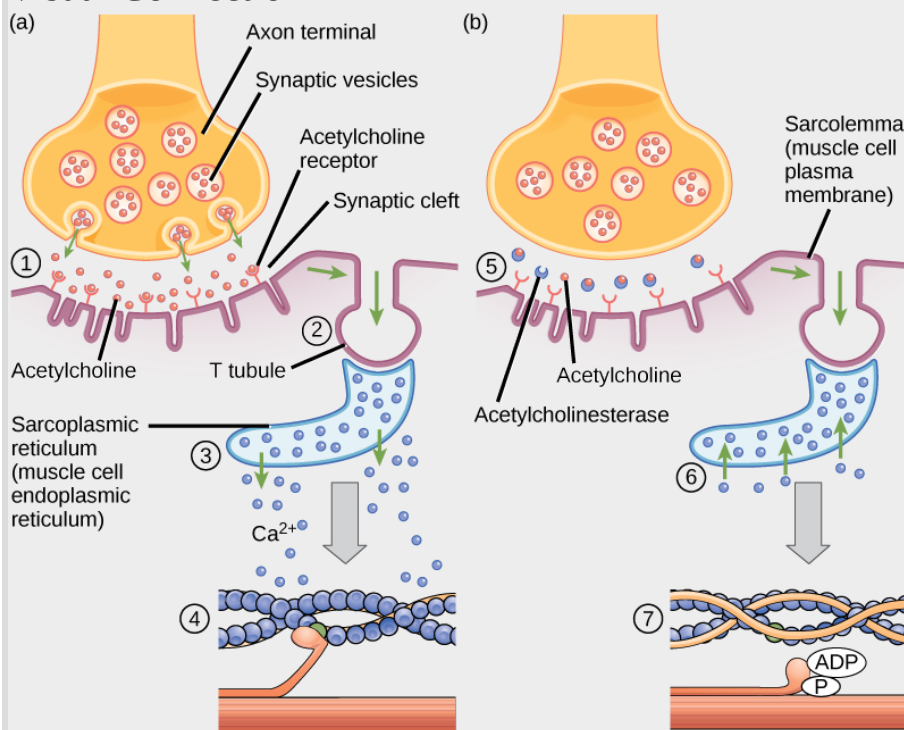
Acetylcholine (ACh) is a neurotransmitter released by motor neurons that binds to receptors in the motor end plate. Neurotransmitter release occurs when an action potential travels down the motor neuron's axon, resulting in altered permeability of the synaptic terminal membrane and an influx of calcium. The Ca^{2+} ions allow synaptic vesicles to move to and bind with the presynaptic membrane (on the neuron), and release neurotransmitter from the vesicles into the synaptic cleft. Once released by the synaptic terminal, ACh diffuses across the synaptic cleft to the motor end plate, where it binds with ACh receptors. As a neurotransmitter binds, these ion channels open, and Na^+ ions cross the membrane into the muscle cell. This reduces the voltage difference between the inside and outside of the cell, which is called

depolarization. As ACh binds at the motor end plate, this depolarization is called an end-plate potential. The depolarization then spreads along the sarcolemma, creating an action potential as sodium channels adjacent to the initial depolarization site sense the change in voltage and open. The action potential moves across the entire cell, creating a wave of depolarization.

ACh is broken down by the enzyme **acetylcholinesterase** (AChE) into acetyl and choline. AChE resides in the synaptic cleft, breaking down ACh so that it does not remain bound to ACh receptors, which would cause unwanted extended muscle contraction ([\[link\]](#)).

Note:

Visual Connection



1. Acetylcholine released from the axon terminal binds to receptors on the sarcolemma.
2. An action potential is generated and travels down the T tubule.
3. Ca^{2+} is released from the sarcoplasmic reticulum in response to the change in voltage.
4. Ca^{2+} binds troponin; Cross-bridges form between actin and myosin.
5. Acetylcholinesterase removes acetylcholine from the synaptic cleft.
6. Ca^{2+} is transported back into the sarcoplasmic reticulum.
7. Tropomyosin binds active sites on actin causing the cross-bridge to detach.

This diagram shows excitation-contraction coupling in a skeletal muscle contraction. The sarcoplasmic reticulum is a specialized endoplasmic reticulum found in muscle cells.

The deadly nerve gas Sarin irreversibly inhibits Acetylcholinesterase. What effect would Sarin have on muscle contraction?

After depolarization, the membrane returns to its resting state. This is called repolarization, during which voltage-gated sodium channels close. Potassium channels continue at 90% conductance. Because the plasma membrane sodium–potassium ATPase always transports ions, the resting state (negatively charged inside relative to the outside) is restored. The period immediately following the transmission of an impulse in a nerve or muscle, in which a neuron or muscle cell regains its ability to transmit another impulse, is called the refractory period. During the refractory period, the membrane cannot generate another action potential. The refractory period allows the voltage-sensitive ion channels to return to their resting configurations. The sodium potassium ATPase continually moves Na^+ back out of the cell and K^+ back into the cell, and the K^+ leaks out leaving negative charge behind. Very quickly, the membrane repolarizes, so that it can again be depolarized.

Control of Muscle Tension

Neural control initiates the formation of actin–myosin cross-bridges, leading to the sarcomere shortening involved in muscle contraction. These contractions extend from the muscle fiber through connective tissue to pull on bones, causing skeletal movement. The pull exerted by a muscle is called tension, and the amount of force created by this tension can vary. This enables the same muscles to move very light objects and very heavy objects. In individual muscle fibers, the amount of tension produced depends on the cross-sectional area of the muscle fiber and the frequency of neural stimulation.

The number of cross-bridges formed between actin and myosin determine the amount of tension that a muscle fiber can produce. Cross-bridges can only form where thick and thin filaments overlap, allowing myosin to bind to actin. If more cross-bridges are formed, more myosin will pull on actin, and more tension will be produced.

The ideal length of a sarcomere during production of maximal tension occurs when thick and thin filaments overlap to the greatest degree. If a sarcomere at rest is stretched past an ideal resting length, thick and thin filaments do not overlap to the greatest degree, and fewer cross-bridges can form. This results in fewer myosin heads pulling on actin, and less tension is produced. As a sarcomere is shortened, the zone of overlap is reduced as the thin filaments reach the H zone, which is composed of myosin tails. Because it is myosin heads that form cross-bridges, actin will not bind to myosin in this zone, reducing the tension produced by this myofiber. If the sarcomere is shortened even more, thin filaments begin to overlap with each other—reducing cross-bridge formation even further, and producing even less tension. Conversely, if the sarcomere is stretched to the point at which thick and thin filaments do not overlap at all, no cross-bridges are formed and no tension is produced. This amount of stretching does not usually occur because accessory proteins, internal sensory nerves, and connective tissue oppose extreme stretching.

The primary variable determining force production is the number of myofibers within the muscle that receive an action potential from the neuron that controls that fiber. When using the biceps to pick up a pencil, the motor cortex of the brain only signals a few neurons of the biceps, and only a few myofibers respond. In vertebrates, each myofiber responds fully if stimulated. When picking up a piano, the motor cortex signals all of the neurons in the biceps and every myofiber participates. This is close to the maximum force the muscle can produce. As mentioned above, increasing the frequency of action potentials (the number of signals per second) can increase the force a bit more, because the tropomyosin is flooded with calcium.

Section Summary

The body contains three types of muscle tissue: skeletal muscle, cardiac muscle, and smooth muscle. Skeletal muscle tissue is composed of sarcomeres, the functional units of muscle tissue. Muscle contraction occurs when sarcomeres shorten, as thick and thin filaments slide past each other, which is called the sliding filament model of muscle contraction. ATP provides the energy for cross-bridge formation and filament sliding. Regulatory proteins, such as troponin and tropomyosin, control cross-bridge formation. Excitation–contraction coupling transduces the electrical signal of the neuron, via acetylcholine, to an electrical signal on the muscle membrane, which initiates force production. The number of muscle fibers contracting determines how much force the whole muscle produces.

Visual Connection Questions

Exercise:

Problem:

[\[link\]](#) Which of the following statements about muscle contraction is true?

- a. The power stroke occurs when ATP is hydrolyzed to ADP and phosphate.
- b. The power stroke occurs when ADP and phosphate dissociate from the myosin head.
- c. The power stroke occurs when ADP and phosphate dissociate from the actin active site.
- d. The power stroke occurs when Ca^{2+} binds the calcium head.

Solution:

[\[link\]](#) B

Exercise:

Problem:

[\[link\]](#) The deadly nerve gas Sarin irreversibly inhibits Acetylcholinesterase. What effect would Sarin have on muscle contraction?

Solution:

[\[link\]](#) In the presence of Sarin, acetylcholine is not removed from the synapse, resulting in continuous stimulation of the muscle plasma membrane. At first, muscle activity is intense and uncontrolled, but the ion gradients dissipate, so electrical signals in the T-tubules are no longer possible. The result is paralysis, leading to death by asphyxiation.

Review Questions**Exercise:****Problem:**

In relaxed muscle, the myosin-binding site on actin is blocked by _____.

- a. titin
- b. troponin
- c. myoglobin
- d. tropomyosin

Solution:

D

Exercise:

Problem: The cell membrane of a muscle fiber is called a _____.

- a. myofibril
- b. sarcolemma
- c. sarcoplasm
- d. myofilament

Solution:

B

Exercise:

Problem:

The muscle relaxes if no new nerve signal arrives. However the neurotransmitter from the previous stimulation is still present in the synapse. The activity of _____ helps to remove this neurotransmitter.

- a. myosin
- b. action potential
- c. tropomyosin
- d. acetylcholinesterase

Solution:

D

Exercise:

Problem:

The ability of a muscle to generate tension immediately after stimulation is dependent on:

- a. myosin interaction with the M line
- b. overlap of myosin and actin
- c. actin attachments to the Z line
- d. none of the above

Solution:

D

Exercise:**Problem:**

Botulinum toxin causes flaccid paralysis of the muscles, and is used for cosmetic purposes under the name Botox. Which of the following is the most likely mechanism of action of Botox?

- a. Botox decreases the production of acetylcholinesterase.
- b. Botox increases calcium release from the sarcoplasmic reticulum.
- c. Botox blocks the ATP binding site in actin.
- d. Botox decreases the release of acetylcholine from motor neurons.

Solution:

D

Critical Thinking Questions**Exercise:****Problem:**

How would muscle contractions be affected if ATP was completely depleted in a muscle fiber?

Solution:

Because ATP is required for myosin to release from actin, muscles would remain rigidly contracted until more ATP was available for the myosin cross-bridge release. This is why dead vertebrates undergo rigor mortis.

Exercise:**Problem:**

What factors contribute to the amount of tension produced in an individual muscle fiber?

Solution:

The cross-sectional area, the length of the muscle fiber at rest, and the frequency of neural stimulation.

Exercise:**Problem:**

What effect will low blood calcium have on neurons? What effect will low blood calcium have on skeletal muscles?

Solution:

Neurons will not be able to release neurotransmitter without calcium. Skeletal muscles have calcium stored and don't need any from the outside.

Exercise:**Problem:**

Skeletal muscles can only produce a mechanical force as they are contracted, but a leg flexes and extends while walking. How can muscles perform this task?

Solution:

Muscles are able to drive locomotion (and other tasks involving opposing motions) because they are paired. When walking, the hamstring muscle contracts first, causing the leg to flex around the knee joint. The quadriceps muscle then contracts (while the hamstring relaxes and extends) to straighten the leg as the foot returns to the ground.

Glossary

actin

globular contractile protein that interacts with myosin for muscle contraction

acetylcholinesterase

(AChE) enzyme that breaks down ACh into acetyl and choline

cardiac muscle tissue

muscle tissue found only in the heart; cardiac contractions pump blood throughout the body and maintain blood pressure

motor end plate

sarcolemma of the muscle fiber that interacts with the neuron

myofibril

long cylindrical structures that lie parallel to the muscle fiber

myofilament

small structures that make up myofibrils

myosin

contractile protein that interacts with actin for muscle contraction

sarcolemma

plasma membrane of a skeletal muscle fiber

sarcomere

functional unit of skeletal muscle

skeletal muscle tissue

forms skeletal muscles, which attach to bones and control locomotion and any movement that can be consciously controlled

smooth muscle tissue

occurs in the walls of hollow organs such as the intestines, stomach, and urinary bladder, and around passages such as the respiratory tract and blood vessels

thick filament

a group of myosin molecules

thin filament

two polymers of actin wound together along with tropomyosin and troponin

tropomyosin

acts to block myosin binding sites on actin molecules, preventing cross-bridge formation and preventing contraction until a muscle receives a neuron signal

troponin

binds to tropomyosin and helps to position it on the actin molecule, and also binds calcium ions

Energy Flow through Ecosystems

By the end of this section, you will be able to do the following:

- Describe how organisms acquire energy in a food web and in associated food chains
- Explain how the efficiency of energy transfers between trophic levels affects ecosystem structure and dynamics
- Discuss trophic levels and how ecological pyramids are used to model them

All living things require energy in one form or another. Energy is required by most complex metabolic pathways (often in the form of adenosine triphosphate, ATP), especially those responsible for building large molecules from smaller compounds, and life itself is an energy-driven process. Living organisms would not be able to assemble macromolecules (proteins, lipids, nucleic acids, and complex carbohydrates) from their monomeric subunits without a constant energy input.

It is important to understand how organisms acquire energy and how that energy is passed from one organism to another through food webs and their constituent food chains. Food webs illustrate how energy flows directionally through ecosystems, including how efficiently organisms acquire it, use it, and how much remains for use by other organisms of the food web.

How Organisms Acquire Energy in a Food Web

Energy is acquired by living things in three ways: photosynthesis, chemosynthesis, and the consumption and digestion of other living or previously living organisms by heterotrophs.

Photosynthetic and chemosynthetic organisms are both grouped into a category known as autotrophs: organisms capable of synthesizing their own food (more specifically, capable of using inorganic carbon as a carbon source). Photosynthetic autotrophs (photoautotrophs) use sunlight as an energy source, whereas chemosynthetic autotrophs (chemoautotrophs) use inorganic molecules as an energy source. Autotrophs are critical for all

ecosystems. Without these organisms, energy would not be available to other living organisms and life itself would not be possible.

Photoautotrophs, such as plants, algae, and photosynthetic bacteria, serve as the energy source for a majority of the world's ecosystems. These ecosystems are often described by grazing food webs. Photoautotrophs harness the solar energy of the sun by converting it to chemical energy in the form of ATP (and NADP). The energy stored in ATP is used to synthesize complex organic molecules, such as glucose.

Chemoautotrophs are primarily bacteria that are found in rare ecosystems where sunlight is not available, such as in those associated with dark caves or hydrothermal vents at the bottom of the ocean ([\[link\]](#)). Many chemoautotrophs in hydrothermal vents use hydrogen sulfide (H_2S), which is released from the vents as a source of chemical energy. This allows chemoautotrophs to synthesize complex organic molecules, such as glucose, for their own energy and in turn supplies energy to the rest of the ecosystem.



Swimming shrimp, a few squat lobsters, and hundreds of vent mussels are seen at a hydrothermal vent at the bottom of the

ocean. As no sunlight penetrates to this depth, the ecosystem is supported by chemoautotrophic bacteria and organic material that sinks from the ocean's surface. This picture was taken in 2006 at the submerged NW Eifuku volcano off the coast of Japan by the National Oceanic and Atmospheric Administration (NOAA). The summit of this highly active volcano lies 1535 m below the surface.

Productivity within Trophic Levels

Productivity within an ecosystem can be defined as the percentage of energy entering the ecosystem incorporated into biomass in a particular trophic level. **Biomass** is the total mass, in a unit area at the time of measurement, of living or previously living organisms within a trophic level. Ecosystems have characteristic amounts of biomass at each trophic level. For example, in the English Channel ecosystem the primary producers account for a biomass of 4 g/m² (grams per square meter), while the primary consumers exhibit a biomass of 21 g/m².

The productivity of the primary producers is especially important in any ecosystem because these organisms bring energy to other living organisms by photoautotrophy or chemoautotrophy. The rate at which photosynthetic primary producers incorporate energy from the sun is called **gross primary productivity**. An example of gross primary productivity is shown in the compartment diagram of energy flow within the Silver Springs aquatic ecosystem as shown ([link](#)). In this ecosystem, the total energy accumulated by the primary producers (gross primary productivity) was shown to be 20,810 kcal/m²/yr.

Because all organisms need to use some of this energy for their own functions (like respiration and resulting metabolic heat loss) scientists often

refer to the net primary productivity of an ecosystem. **Net primary productivity** is the energy that remains in the primary producers after accounting for the organisms' respiration and heat loss. The net productivity is then available to the primary consumers at the next trophic level. In our Silver Springs example, 13,187 of the 20,810 kcal/m²/yr were used for respiration or were lost as heat, leaving 7,633 kcal/m²/yr of energy for use by the primary consumers.

Ecological Efficiency: The Transfer of Energy between Trophic Levels

As illustrated in ([link](#)), as energy flows from primary producers through the various trophic levels, the ecosystem loses large amounts of energy. The main reason for this loss is the second law of thermodynamics, which states that whenever energy is converted from one form to another, there is a tendency toward disorder (entropy) in the system. In biologic systems, this energy takes the form of metabolic heat, which is lost when the organisms consume other organisms. In the Silver Springs ecosystem example ([link](#)), we see that the primary consumers produced 1103 kcal/m²/yr from the 7618 kcal/m²/yr of energy available to them from the primary producers. The measurement of energy transfer efficiency between two successive trophic levels is termed the **trophic level transfer efficiency (TLTE)** and is defined by the formula:

Equation:

$$\text{TLTE} = \frac{\text{production at present trophic level}}{\text{production at previous trophic level}} \times 100$$

In Silver Springs, the TLTE between the first two trophic levels was approximately 14.8 percent. The low efficiency of energy transfer between trophic levels is usually the major factor that limits the length of food chains observed in a food web. The fact is, after four to six energy transfers, there is not enough energy left to support another trophic level. In the Lake Ontario example shown in ([link](#)), only three energy transfers occurred between the primary producer, (green algae), and the apex consumer (Chinook salmon).

Ecologists have many different methods of measuring energy transfers within ecosystems. Measurement difficulty depends on the complexity of the ecosystem and how much access scientists have to observe the ecosystem. In other words, some ecosystems are more difficult to study than others, and sometimes the quantification of energy transfers has to be estimated.

Other parameters are important in characterizing energy flow within an ecosystem. **Net production efficiency (NPE)** allows ecologists to quantify how efficiently organisms of a particular trophic level incorporate the energy they receive into biomass; it is calculated using the following formula:

Equation:

$$\text{NPE} = \frac{\text{net consumer productivity}}{\text{assimilation}} \times 100$$

Net consumer productivity is the energy content available to the organisms of the next trophic level. **Assimilation** is the biomass (energy content generated per unit area) of the present trophic level after accounting for the energy lost due to incomplete ingestion of food, energy used for respiration, and energy lost as waste. Incomplete ingestion refers to the fact that some consumers eat only a part of their food. For example, when a lion kills an antelope, it will eat everything except the hide and bones. The lion is missing the energy-rich bone marrow inside the bone, so the lion does not make use of all the calories its prey could provide.

Thus, NPE measures how efficiently each trophic level uses and incorporates the energy from its food into biomass to fuel the next trophic level. In general, cold-blooded animals (ectotherms), such as invertebrates, fish, amphibians, and reptiles, use less of the energy they obtain for respiration and heat than warm-blooded animals (endotherms), such as birds and mammals. The extra heat generated in endotherms, although an advantage in terms of the activity of these organisms in colder environments, is a major disadvantage in terms of NPE. Therefore, many endotherms have to eat more often than ectotherms to get the energy they need for survival. In general, NPE for ectotherms is an order of magnitude

(10x) higher than for endotherms. For example, the NPE for a caterpillar eating leaves has been measured at 18 percent, whereas the NPE for a squirrel eating acorns may be as low as 1.6 percent.

The inefficiency of energy use by warm-blooded animals has broad implications for the world's food supply. It is widely accepted that the meat industry uses large amounts of crops to feed livestock, and because the NPE is low, much of the energy from animal feed is lost. For example, it costs about \$0.01 to produce 1000 dietary calories (kcal) of corn or soybeans, but approximately \$0.19 to produce a similar number of calories growing cattle for beef consumption. The same energy content of milk from cattle is also costly, at approximately \$0.16 per 1000 kcal. Much of this difference is due to the low NPE of cattle. Thus, there has been a growing movement worldwide to promote the consumption of nonmeat and nondairy foods so that less energy is wasted feeding animals for the meat industry.

Modeling Ecosystems Energy Flow: Ecological Pyramids

The structure of ecosystems can be visualized with ecological pyramids, which were first described by the pioneering studies of Charles Elton in the 1920s. **Ecological pyramids** show the relative amounts of various parameters (such as number of organisms, energy, and biomass) across trophic levels.

Pyramids of numbers can be either upright or inverted, depending on the ecosystem. As shown in [\[link\]](#), typical grassland during the summer has a base of many plants, and the numbers of organisms decrease at each trophic level. However, during the summer in a temperate forest, the base of the pyramid consists of few trees compared with the number of primary consumers, mostly insects. Because trees are large, they have great photosynthetic capability, and dominate other plants in this ecosystem to obtain sunlight. Even in smaller numbers, primary producers in forests are still capable of supporting other trophic levels.

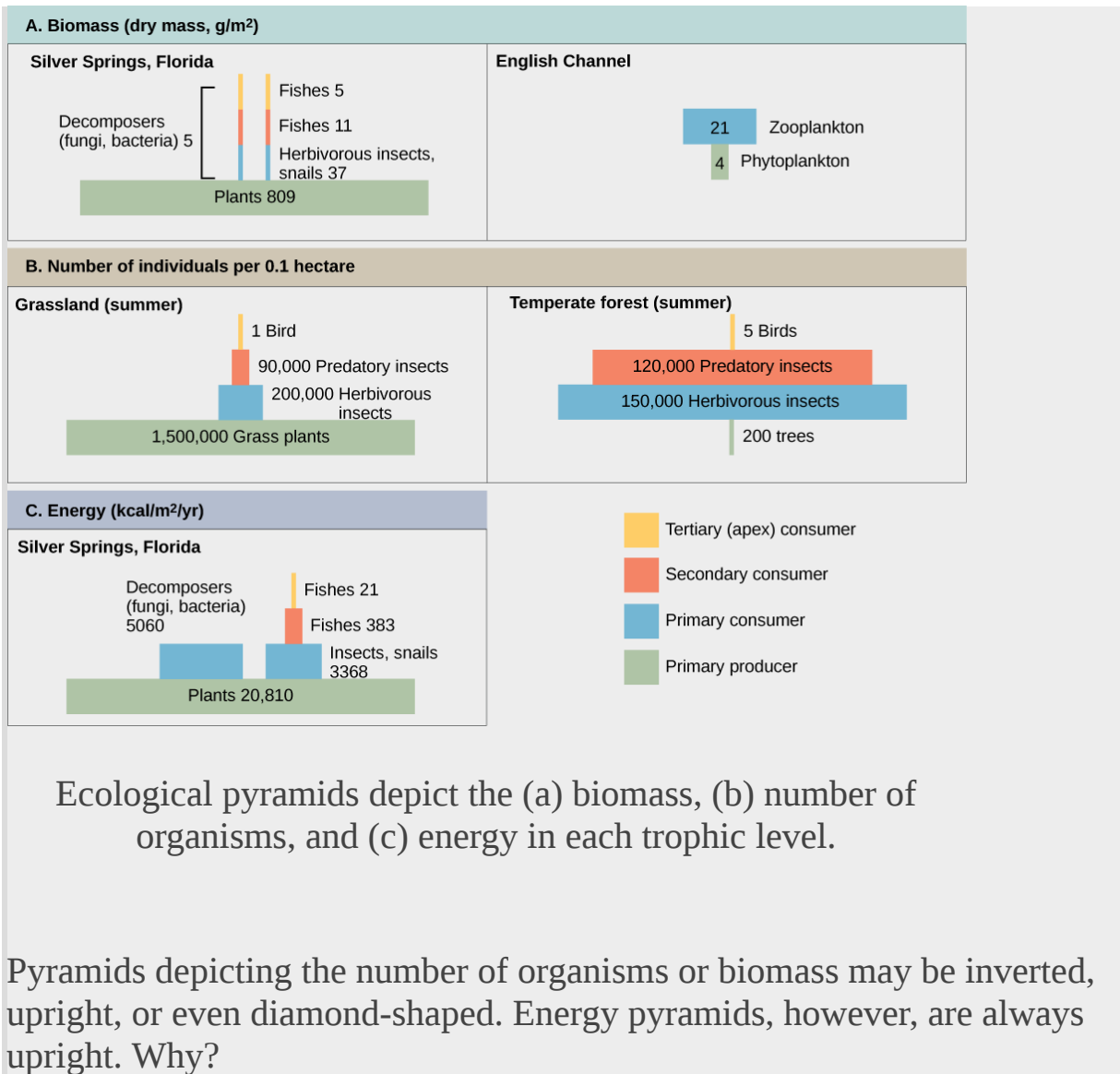
Another way to visualize ecosystem structure is with pyramids of biomass. This pyramid measures the amount of energy converted into living tissue at the different trophic levels. Using the Silver Springs ecosystem example,

this data exhibits an upright biomass pyramid ([\[link\]](#)), whereas the pyramid from the English Channel example is inverted. The plants (primary producers) of the Silver Springs ecosystem make up a large percentage of the biomass found there. However, the phytoplankton in the English Channel example make up less biomass than the primary consumers, the zooplankton. As with inverted pyramids of numbers, this inverted pyramid is not due to a lack of productivity from the primary producers, but results from the high turnover rate of the phytoplankton. The phytoplankton are consumed rapidly by the primary consumers, thus, minimizing their biomass at any particular point in time. However, phytoplankton reproduce quickly, thus they are able to support the rest of the ecosystem.

Pyramid ecosystem modeling can also be used to show energy flow through the trophic levels. Notice that these numbers are the same as those used in the energy flow compartment diagram in ([\[link\]](#)). Pyramids of energy are always upright, and an ecosystem without sufficient primary productivity cannot be supported. All types of ecological pyramids are useful for characterizing ecosystem structure. However, in the study of energy flow through the ecosystem, pyramids of energy are the most consistent and representative models of ecosystem structure ([\[link\]](#)).

Note:

Visual Connection

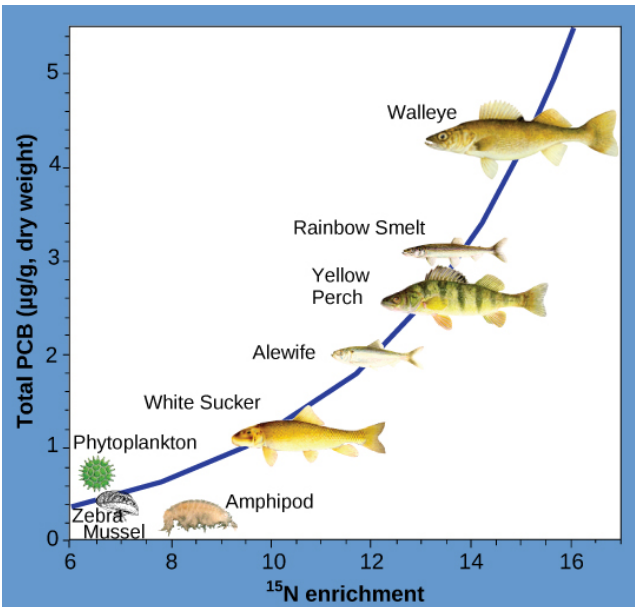


Consequences of Food Webs: Biological Magnification

One of the most important environmental consequences of ecosystem dynamics is biomagnification. **Biomagnification** is the increasing concentration of persistent, toxic substances in organisms at each trophic level, from the primary producers to the apex consumers. Many substances have been shown to bioaccumulate, including the pesticide **dichlorodiphenyltrichloroethane (DDT)**, which was described in the 1960s bestseller, *Silent Spring*, by Rachel Carson. DDT was a commonly used

pesticide before its dangers became known. In some aquatic ecosystems, organisms from each trophic level consumed many organisms of the lower level, which caused DDT to increase in birds (apex consumers) that ate fish. Thus, the birds accumulated sufficient amounts of DDT to cause fragility in their eggshells. This effect increased egg breakage during nesting and was shown to have adverse effects on these bird populations. The use of DDT was banned in the United States in the 1970s.

Other substances that biomagnify are polychlorinated biphenyls (PCBs), which were used in coolant liquids in the United States until their use was banned in 1979, and heavy metals, such as mercury, lead, and cadmium. These substances were best studied in aquatic ecosystems, where fish species at different trophic levels accumulate toxic substances brought through the ecosystem by the primary producers. As illustrated in a study performed by the National Oceanic and Atmospheric Administration (NOAA) in the Saginaw Bay of Lake Huron ([\[link\]](#)), PCB concentrations increased from the ecosystem's primary producers (phytoplankton) through the different trophic levels of fish species. The apex consumer (walleye) has more than four times the amount of PCBs compared to phytoplankton. Also, based on results from other studies, birds that eat these fish may have PCB levels at least one order of magnitude higher than those found in the lake fish.



This chart shows the PCB concentrations found at the various trophic levels in the Saginaw Bay ecosystem of Lake Huron. Numbers on the x-axis reflect enrichment with heavy isotopes of nitrogen (^{15}N), which is a marker for increasing trophic level. Notice that the fish in the higher trophic levels accumulate more PCBs than those in lower trophic levels. (credit: Patricia Van Hoof, NOAA, GLERL)

Other concerns have been raised by the accumulation of heavy metals, such as mercury and cadmium, in certain types of seafood. The United States Environmental Protection Agency (EPA) recommends that pregnant women and young children should not consume any swordfish, shark, king mackerel, or tilefish because of their high mercury content. These individuals are advised to eat fish low in mercury: salmon, tilapia, shrimp,

pollock, and catfish. Biomagnification is a good example of how ecosystem dynamics can affect our everyday lives, even influencing the food we eat.

Section Summary

Organisms in an ecosystem acquire energy in a variety of ways, which is transferred between trophic levels as the energy flows from the bottom to the top of the food web, with energy being lost at each transfer. The efficiency of these transfers is important for understanding the different behaviors and eating habits of warm-blooded versus cold-blooded animals. Modeling of ecosystem energy is best done with ecological pyramids of energy, although other ecological pyramids provide other vital information about ecosystem structure.

Visual Connection Questions

Exercise:

Problem:

[\[link\]](#) Pyramids depicting the number of organisms or biomass may be inverted, upright, or even diamond-shaped. Energy pyramids, however, are always upright. Why?

Solution:

[\[link\]](#) Pyramids of organisms may be inverted or diamond-shaped because a large organism, such as a tree, can sustain many smaller organisms. Likewise, a low biomass of organisms can sustain a larger biomass at the next trophic level because the organisms reproduce rapidly and thus supply continuous nourishment. Energy pyramids, however, must always be upright because of the laws of thermodynamics. The first law of thermodynamics states that energy can neither be created nor destroyed; thus, each trophic level must acquire energy from the trophic level below. The second law of thermodynamics states that, during the transfer of energy, some energy

is always lost as heat; thus, less energy is available at each higher trophic level.

Review Questions

Exercise:

Problem:

The weight of living organisms in an ecosystem at a particular point in time is called:

- a. energy
- b. production
- c. entropy
- d. biomass

Solution:

D

Exercise:

Problem:

Which term describes the process whereby toxic substances increase along trophic levels of an ecosystem?

- a. biomassification
- b. biomagnification
- c. bioentropy
- d. heterotrophy

Solution:

B

Exercise:

Problem:

Organisms that can make their own food using inorganic molecules are called:

- a. autotrophs
- b. heterotrophs
- c. photoautotrophs
- d. chemoautotrophs

Solution:

D

Exercise:**Problem:**

In the English Channel ecosystem, the number of primary producers is smaller than the number of primary consumers because_____.

- a. the apex consumers have a low turnover rate
- b. the primary producers have a low turnover rate
- c. the primary producers have a high turnover rate
- d. the primary consumers have a high turnover rate

Solution:

C

Exercise:**Problem:**

What law of chemistry determines how much energy can be transferred when it is converted from one form to another?

- a. the first law of thermodynamics

- b. the second law of thermodynamics
- c. the conservation of matter
- d. the conservation of energy

Solution:

B

Exercise:

Problem:

The mussels that live at the NW Eifuku volcano are examples of _____.

- a. chemoautotrophs
- b. photoautotrophs
- c. apex predators
- d. primary consumers

Solution:

D

Critical Thinking Questions

Exercise:

Problem:

Compare the three types of ecological pyramids and how well they describe ecosystem structure. Identify which ones can be inverted and give an example of an inverted pyramid for each.

Solution:

Pyramids of numbers display the number of individual organisms on each trophic level. These pyramids can be either upright or inverted, depending on the number of the organisms. Pyramids of biomass display the weight of organisms at each level. Inverted pyramids of biomass can occur when the primary producer has a high turnover rate. Pyramids of energy are usually upright and are the best representation of energy flow and ecosystem structure.

Exercise:

Problem:

How does the amount of food a warm-blooded animal (endotherm) eats relate to its net production efficiency (NPE)?

Solution:

NPE measures the rate at which one trophic level can use and make biomass from what it attained in the previous level, taking into account respiration, defecation, and heat loss. Endotherms have high metabolism and generate a lot of body heat. Although this gives them advantages in their activity level in colder temperatures, these organisms are 10 times less efficient at harnessing the energy from the food they eat compared with cold-blooded animals, and thus have to eat more and more often.

Exercise:

Problem:

A study uses an inverted pyramid to demonstrate the relationship between sharks, their aquatic prey, and phytoplankton in an ocean region. What type of pyramid must be used? What does this convey to readers about predation in the area?

Solution:

An inverted ecological pyramid describing the relationship between the three groups must be a biomass pyramid. This model suggests that the area is subject to heavy predation, with the prey species feeding

heavily on the phytoplankton, and in turn being consumed by the sharks.

Exercise:

Problem:

Describe what a pyramid of numbers would look like if an ecologist models the relationship between bird parasites, blue jays, and oak trees in a hectare. Does this match the energy flow pyramid?

Solution:

In this ecological model, the oak trees (producers) would be at the bottom, the blue jays would be in the middle level (primary consumer of acorns), and the parasites would be at the top level (secondary consumer). However, the pyramid would be inverted since each bird could support several parasites, and each tree could support several birds. This pyramid would appear to be the opposite of the energy flow pyramid.

Glossary

assimilation

biomass consumed and assimilated from the previous trophic level after accounting for the energy lost due to incomplete ingestion of food, energy used for respiration, and energy lost as waste

biomagnification

increasing concentrations of persistent, toxic substances in organisms at each trophic level, from the primary producers to the apex consumers

biomass

total weight, at the time of measurement, of living or previously living organisms in a unit area within a trophic level

chemoautotroph

organism capable of synthesizing its own food using energy from inorganic molecules

ecological pyramid

(also, Eltonian pyramid) graphical representation of different trophic levels in an ecosystem based of organism numbers, biomass, or energy content

gross primary productivity

rate at which photosynthetic primary producers incorporate energy from the sun

net consumer productivity

energy content available to the organisms of the next trophic level

net primary productivity

energy that remains in the primary producers after accounting for the organisms' respiration and heat loss

net production efficiency (NPE)

measure of the ability of a trophic level to convert the energy it receives from the previous trophic level into biomass

trophic level transfer efficiency (TLTE)

energy transfer efficiency between two successive trophic levels

Biogeochemical Cycles

By the end of this section, you will be able to do the following:

- Discuss the biogeochemical cycles of water, carbon, nitrogen, phosphorus, and sulfur
- Explain how human activities have impacted these cycles and the potential consequences for Earth

Energy flows directionally through ecosystems, entering as sunlight (or inorganic molecules for chemoautotrophs) and leaving as heat during the many transfers between trophic levels. However, the matter that makes up living organisms is conserved and recycled. The six most common elements associated with organic molecules—carbon, nitrogen, hydrogen, oxygen, phosphorus, and sulfur—take a variety of chemical forms and may exist for long periods in the atmosphere, on land, in water, or beneath the Earth's surface. Geologic processes, such as weathering, erosion, water drainage, and the subduction of the continental plates, all play a role in this recycling of materials. Because geology and chemistry have major roles in the study of this process, the recycling of inorganic matter between living organisms and their environment is called a **biogeochemical cycle**.

Water contains hydrogen and oxygen, which is essential to all living processes. The **hydrosphere** is the area of the Earth where water movement and storage occurs. On or beneath the surface, water occurs in liquid or solid form in rivers, lakes, oceans, groundwater, polar ice caps, and glaciers. And it occurs as water vapor in the atmosphere. Carbon is found in all organic macromolecules and is an important constituent of fossil fuels. Nitrogen is a major component of our nucleic acids and proteins and is critical to human agriculture. Phosphorus, a major component of nucleic acid (along with nitrogen), is one of the main ingredients in artificial fertilizers used in agriculture and their associated environmental impacts on our surface water. Sulfur is critical to the 3-D folding of proteins, such as in disulfide binding.

The cycling of these elements is interconnected. For example, the movement of water is critical for the leaching of nitrogen and phosphate into rivers, lakes, and oceans. Furthermore, the ocean itself is a major reservoir for carbon. Thus, mineral nutrients are cycled, either rapidly or

slowly, through the entire biosphere, from one living organism to another, and between the biotic and abiotic world.

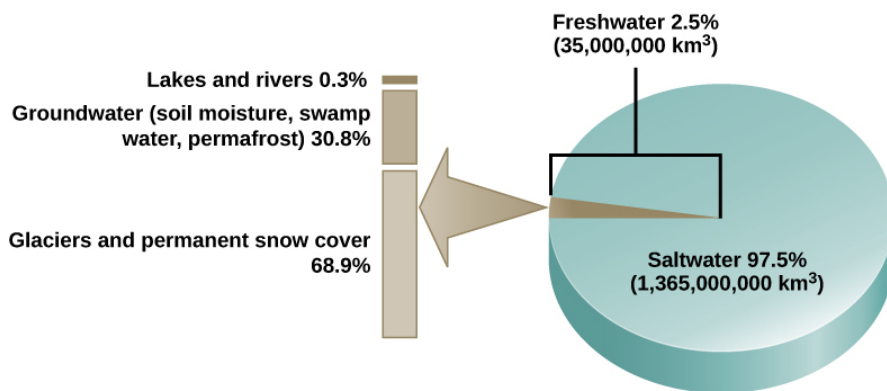
Note:

Link to Learning

Head to this [website](#) to learn more about biogeochemical cycles.

The Water (Hydrologic) Cycle

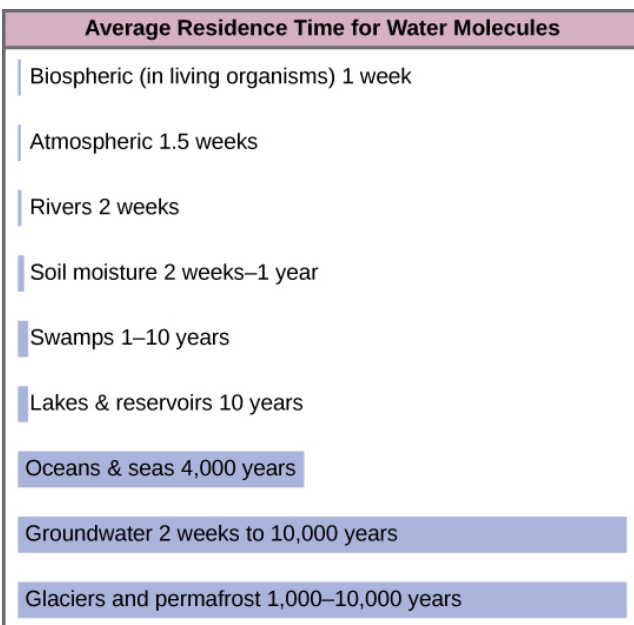
Water is the basis of all living processes on Earth. When examining the stores of water on Earth, 97.5 percent of it is non-potable salt water ([link](#)). Of the remaining water, 99 percent is locked underground as water or as ice. Thus, less than 1 percent of fresh water is easily accessible from lakes and rivers. Many living things, such as plants, animals, and fungi, are dependent on that small amount of fresh surface water, a lack of which can have massive effects on ecosystem dynamics. To be successful, organisms must adapt to fluctuating water supplies. Humans, of course, have developed technologies to increase water availability, such as digging wells to harvest groundwater, storing rainwater, and using desalination to obtain drinkable water from the ocean.



Only 2.5 percent of water on Earth is fresh water,

and less than 1 percent of fresh water is easily accessible to living things.

Water cycling is extremely important to ecosystem dynamics. Water has a major influence on climate and, thus, on the environments of ecosystems. Most of the water on Earth is stored for long periods in the oceans, underground, and as ice. [\[link\]](#) illustrates the average time that an individual water molecule may spend in the Earth's major water reservoirs. **Residence time** is a measure of the average time an individual water molecule stays in a particular reservoir.



This graph shows the average residence time for water molecules in the Earth's water reservoirs.

There are various processes that occur during the cycling of water, shown in [\[link\]](#). These processes include the following:

- evaporation/sublimation
- condensation/precipitation
- subsurface water flow
- surface runoff/snowmelt
- streamflow

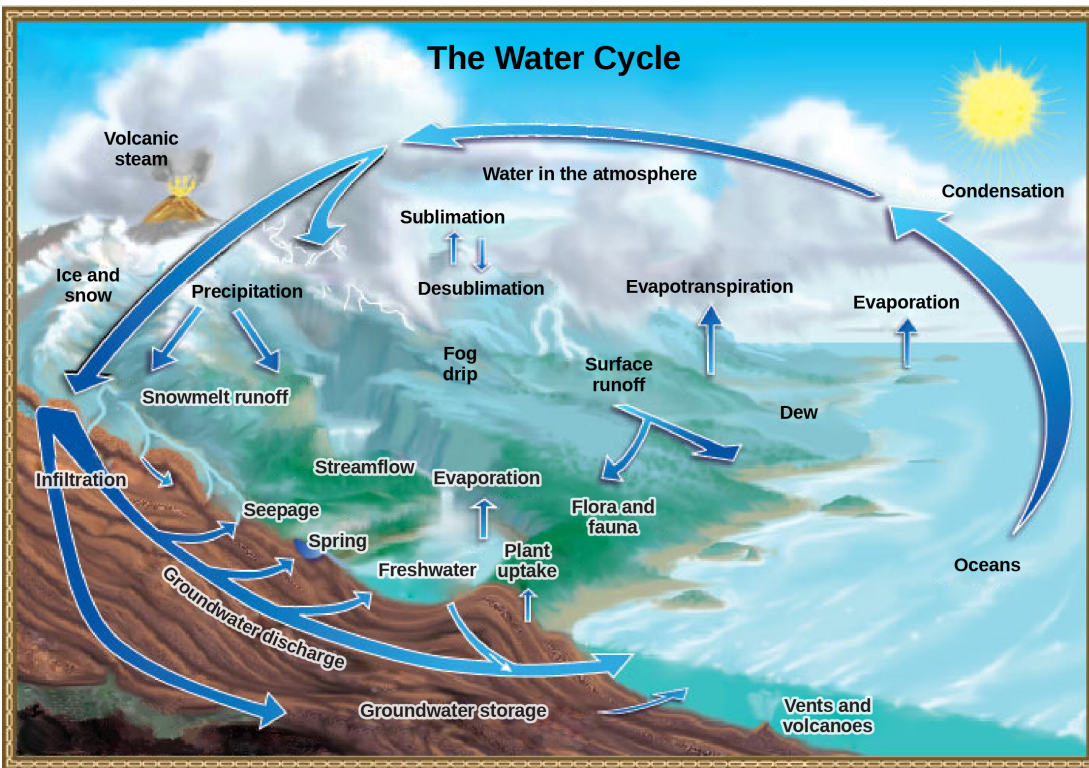
The water cycle is driven by the sun's energy as it warms the oceans and other surface waters. This leads to the evaporation (water to water vapor) of liquid surface water and the sublimation (ice to water vapor) of frozen water, which deposits large amounts of water vapor into the atmosphere. Over time, this water vapor condenses into clouds as liquid or frozen droplets and is eventually followed by precipitation (rain or snow), which returns water to the Earth's surface. Rain eventually permeates into the ground, where it may evaporate again if it is near the surface, flow beneath the surface, or be stored for long periods. More easily observed is surface runoff: the flow of fresh water either from rain or melting ice. Runoff can then make its way through streams and lakes to the oceans or flow directly to the oceans themselves.

Note:

Link to Learning

Head to this [website](#) to learn more about the world's fresh water supply.

Rain and surface runoff are major ways in which minerals, including carbon, nitrogen, phosphorus, and sulfur, are cycled from land to water. The environmental effects of runoff will be discussed later as these cycles are described.

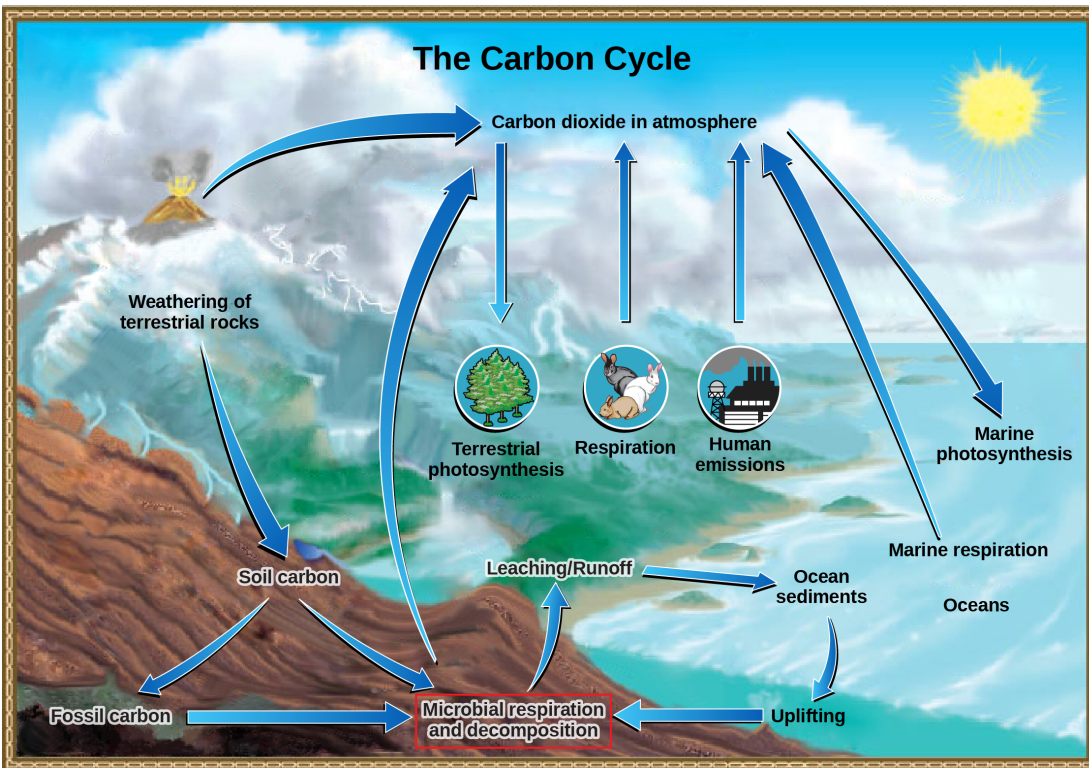


Water from the land and oceans enters the atmosphere by evaporation or sublimation, where it condenses into clouds and falls as rain or snow. Precipitated water may enter freshwater bodies or infiltrate the soil. The cycle is complete when surface or groundwater reenters the ocean. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

The Carbon Cycle

Carbon is the second most abundant element in living organisms. Carbon is present in all organic molecules, and its role in the structure of macromolecules is of primary importance to living organisms.

The carbon cycle is most easily studied as two interconnected sub-cycles: one dealing with rapid carbon exchange among living organisms and the other dealing with the long-term cycling of carbon through geologic processes. The entire carbon cycle is shown in [\[link\]](#).



Carbon dioxide gas exists in the atmosphere and is dissolved in water. Photosynthesis converts carbon dioxide gas to organic carbon, and respiration cycles the organic carbon back into carbon dioxide gas. Long-term storage of organic carbon occurs when matter from living organisms is buried deep underground and becomes fossilized. Volcanic activity and, more recently, human emissions, bring this stored carbon back into the carbon cycle. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Note:

Link to Learning

Click this [link](#) to read information about the United States Carbon Cycle Science Program.

The Biological Carbon Cycle

Living organisms are connected in many ways, even between ecosystems. A good example of this connection is the exchange of carbon between autotrophs and heterotrophs within and between ecosystems by way of atmospheric carbon dioxide. Carbon dioxide is the basic building block that most autotrophs use to build multicarbon, high energy compounds, such as glucose. The energy harnessed from the sun is used by these organisms to form the covalent bonds that link carbon atoms together. These chemical bonds thereby store this energy for later use in the process of respiration. Most terrestrial autotrophs obtain their carbon dioxide directly from the atmosphere, while marine autotrophs acquire it in the dissolved form (carbonic acid, H_2CO_3^-). However carbon dioxide is acquired, a by-product of the process is oxygen. The photosynthetic organisms are responsible for depositing approximately 21 percent oxygen content of the atmosphere that we observe today.

Heterotrophs and autotrophs are partners in biological carbon exchange (especially the primary consumers, largely herbivores). Heterotrophs acquire the high-energy carbon compounds from the autotrophs by consuming them, and breaking them down by respiration to obtain cellular energy, such as ATP. The most efficient type of respiration, aerobic respiration, requires oxygen obtained from the atmosphere or dissolved in water. Thus, there is a constant exchange of oxygen and carbon dioxide between the autotrophs (which need the carbon) and the heterotrophs (which need the oxygen). Gas exchange through the atmosphere and water is one way that the carbon cycle connects all living organisms on Earth.

The Biogeochemical Carbon Cycle

The movement of carbon through the land, water, and air is complex, and in many cases, it occurs much more slowly geologically than as seen between living organisms. Carbon is stored for long periods in what are known as carbon reservoirs, which include the atmosphere, bodies of liquid water (mostly oceans), ocean sediment, soil, land sediments (including fossil fuels), and the Earth's interior.

As stated, the atmosphere is a major reservoir of carbon in the form of carbon dioxide and is essential to the process of photosynthesis. The level of carbon dioxide in the atmosphere is greatly influenced by the reservoir of carbon in the oceans. The exchange of carbon between the atmosphere and water reservoirs influences how much carbon is found in each location, and each one affects the other reciprocally. Carbon dioxide (CO₂) from the atmosphere dissolves in water and combines with water molecules to form carbonic acid, and then it ionizes to carbonate and bicarbonate ions ([\[link\]](#))

Step 1: CO ₂ (atmospheric) $\rightleftharpoons$ CO ₂ (dissolved)
Step 2: CO ₂ (dissolved) + H ₂ O $\rightleftharpoons$ H ₂ CO ₃ (carbonic acid)
Step 3: H ₂ CO ₃ $\rightleftharpoons$ H ⁺ + HCO ₃ ⁻ (bicarbonate ion)
Step 4: HCO ₃ ⁻ $\rightleftharpoons$ H ⁺ + CO ₃ ²⁻ (carbonate ion)

Carbon dioxide reacts with water to form bicarbonate and carbonate ions.

The equilibrium coefficients are such that more than 90 percent of the carbon in the ocean is found as bicarbonate ions. Some of these ions combine with seawater calcium to form calcium carbonate (CaCO₃), a major component of marine organism shells. These organisms eventually form sediments on the ocean floor. Over geologic time, the calcium carbonate forms limestone, which comprises the largest carbon reservoir on Earth.

On land, carbon is stored in soil as a result of the decomposition of living organisms (by decomposers) or from weathering of terrestrial rock and minerals. This carbon can be leached into the water reservoirs by surface runoff. Deeper underground, on land and at sea, are fossil fuels: the anaerobically decomposed remains of plants that take millions of years to form. Fossil fuels are considered a nonrenewable resource because their use far exceeds their rate of formation. A **nonrenewable resource**, such as

fossil fuel, is either regenerated very slowly or not at all. Another way for carbon to enter the atmosphere is from land (including land beneath the surface of the ocean) by the eruption of volcanoes and other geothermal systems. Carbon sediments from the ocean floor are taken deep within the Earth by the process of **subduction**: the movement of one tectonic plate beneath another. Carbon is released as carbon dioxide when a volcano erupts or from volcanic hydrothermal vents.

Humans contribute to atmospheric carbon by the burning of fossil fuels and other materials. Since the Industrial Revolution, humans have significantly increased the release of carbon and carbon compounds, which has in turn affected the climate and overall environment.

Animal husbandry by humans also increases atmospheric carbon. The large numbers of land animals raised to feed the Earth's growing population results in increased carbon dioxide levels in the atmosphere due to farming practices and respiration and methane production. This is another example of how human activity indirectly affects biogeochemical cycles in a significant way. Although much of the debate about the future effects of increasing atmospheric carbon on climate change focuses on fossil fuels, scientists take natural processes, such as volcanoes and respiration, into account as they model and predict the future impact of this increase.

The Nitrogen Cycle

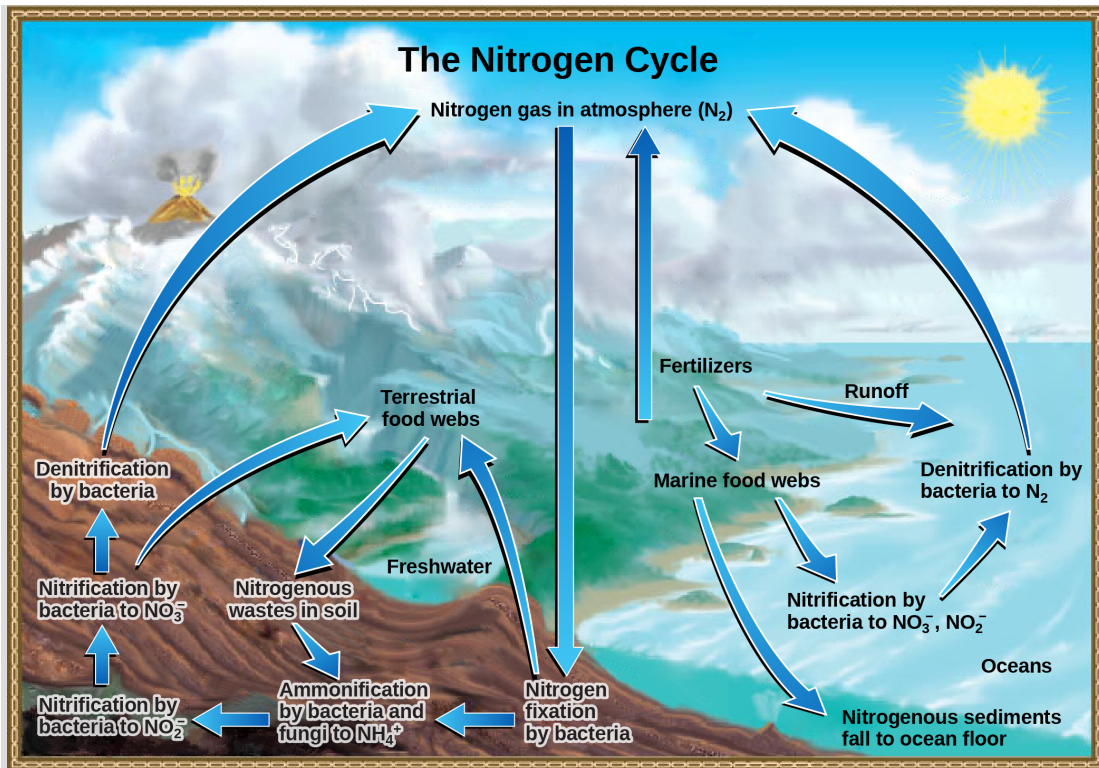
Getting nitrogen into the living world is difficult. Plants and phytoplankton are not equipped to incorporate nitrogen from the atmosphere (which exists as tightly bonded, triple covalent N_2) even though this molecule comprises approximately 78 percent of the atmosphere. Nitrogen enters the living world via free-living and symbiotic bacteria, which incorporate nitrogen into their macromolecules through nitrogen fixation (conversion of N_2). Cyanobacteria live in most aquatic ecosystems where sunlight is present; they play a key role in nitrogen fixation. Cyanobacteria are able to use inorganic sources of nitrogen to “fix” nitrogen. *Rhizobium* bacteria live symbiotically in the root nodules of legumes (such as peas, beans, and peanuts) and provide them with the organic nitrogen they need. (For example, gardeners often grow peas both for their produce and to naturally

add nitrogen to the soil. This practice goes back to ancient times, even if the science has only been recently understood.) Free-living bacteria, such as *Azotobacter*, are also important nitrogen fixers.

Organic nitrogen is especially important to the study of ecosystem dynamics since many ecosystem processes, such as primary production and decomposition, are limited by the available supply of nitrogen. As shown in [\[link\]](#), the nitrogen that enters living systems by nitrogen fixation is successively converted from organic nitrogen back into nitrogen gas by bacteria. This process occurs in three steps in terrestrial systems: ammonification, nitrification, and denitrification. First, the ammonification process converts nitrogenous waste from living animals or from the remains of dead animals into ammonium (NH_4^+) by certain bacteria and fungi. Second, the ammonium is converted to nitrites (NO_2^-) by nitrifying bacteria, such as *Nitrosomonas*, through nitrification. Subsequently, nitrites are converted to nitrates (NO_3^-) by similar organisms. Third, the process of denitrification occurs, whereby bacteria, such as *Pseudomonas* and *Clostridium*, convert the nitrates into nitrogen gas, allowing it to reenter the atmosphere.

Note:

Visual Connection



Nitrogen enters the living world from the atmosphere via nitrogen-fixing bacteria. This nitrogen and nitrogenous waste from animals is then processed back into gaseous nitrogen by soil bacteria, which also supply terrestrial food webs with the organic nitrogen they need. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

Which of the following statements about the nitrogen cycle is false?

- Ammonification converts organic nitrogenous matter from living organisms into ammonium (NH_4^+).
- Denitrification by bacteria converts nitrates (NO_3^-) to nitrogen gas (N_2).
- Nitrification by bacteria converts nitrates (NO_3^-) to nitrites (NO_2^-).
- Nitrogen fixing bacteria convert nitrogen gas (N_2) into organic compounds.

Human activity can release nitrogen into the environment by two primary means: the combustion of fossil fuels, which releases different nitrogen oxides, and by the use of artificial fertilizers in agriculture, which are then washed into lakes, streams, and rivers by surface runoff. Atmospheric nitrogen is associated with several effects on Earth's ecosystems including the production of acid rain (as nitric acid, HNO_3) and greenhouse gas (as nitrous oxide, N_2O) potentially causing climate change. A major effect from fertilizer runoff is saltwater and freshwater **eutrophication**, a process whereby nutrient runoff causes the excess growth of microorganisms, depleting dissolved oxygen levels and killing ecosystem fauna.

A similar process occurs in the marine nitrogen cycle, where the ammonification, nitrification, and denitrification processes are performed by marine bacteria. Some of this nitrogen falls to the ocean floor as sediment, which can then be moved to land in geologic time by uplift of the Earth's surface and thereby incorporated into terrestrial rock. Although the movement of nitrogen from rock directly into living systems has been traditionally seen as insignificant compared with nitrogen fixed from the atmosphere, a recent study showed that this process may indeed be significant and should be included in any study of the global nitrogen cycle. [\[footnote\]](#)

Scott L. Morford, Benjamin Z. Houlton, and Randy A. Dahlgren, "Increased Forest Ecosystem Carbon and Nitrogen Storage from Nitrogen Rich Bedrock," *Nature* 477, no. 7362 (2011): 78–81.

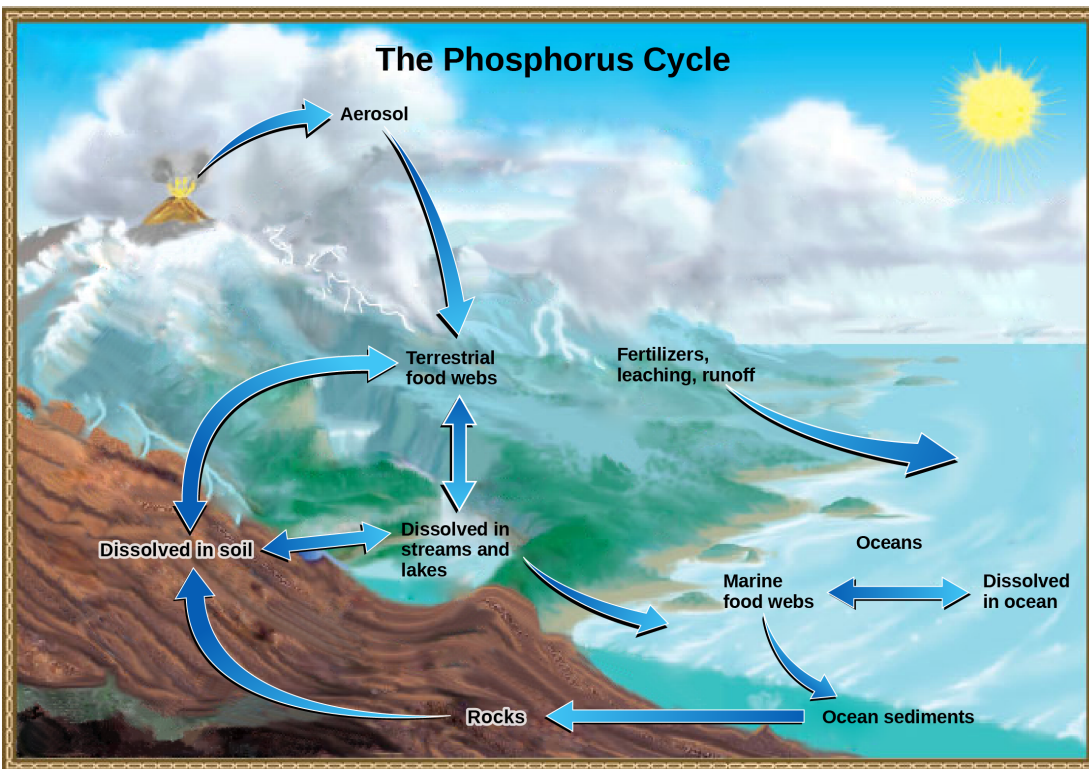
The Phosphorus Cycle

Phosphorus is an essential nutrient for living processes; it is a major component of nucleic acid and phospholipids, and, as calcium phosphate, makes up the supportive components of our bones. Phosphorus is often the limiting nutrient (necessary for growth) in aquatic ecosystems ([\[link\]](#)).

Phosphorus occurs in nature as the phosphate ion (PO_4^{3-}). In addition to phosphate runoff as a result of human activity, natural surface runoff occurs when it is leached from phosphate-containing rock by weathering, thus sending phosphates into rivers, lakes, and the ocean. This rock has its origins in the ocean. Phosphate-containing ocean sediments form primarily

from the bodies of ocean organisms and from their excretions. However, in remote regions, volcanic ash, aerosols, and mineral dust may also be significant phosphate sources. This sediment then is moved to land over geologic time by the uplifting of areas of the Earth's surface.

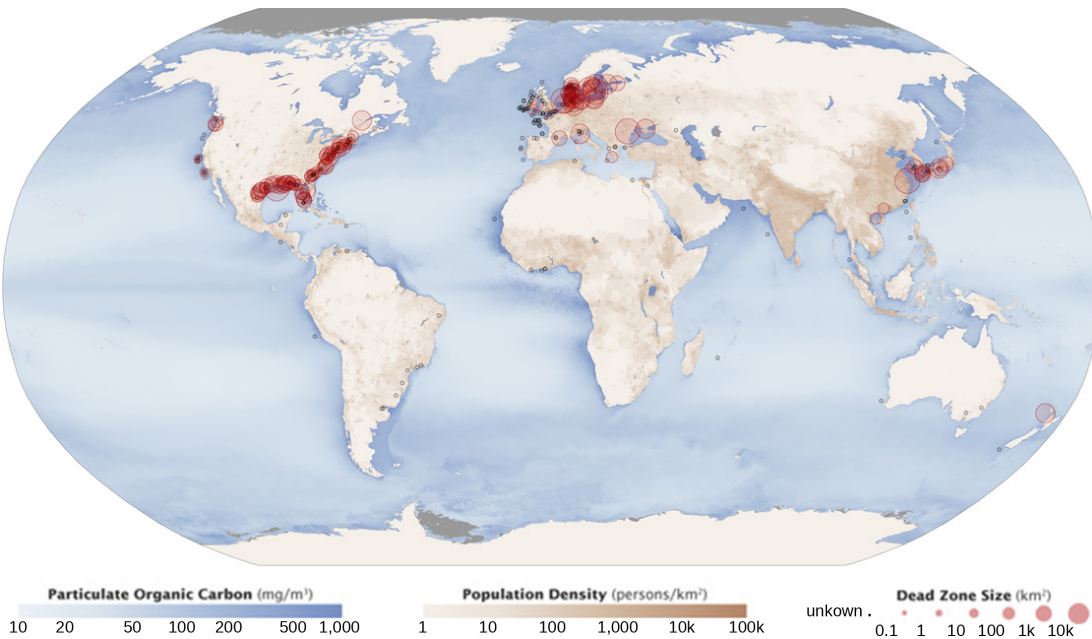
Phosphorus is also reciprocally exchanged between phosphate dissolved in the ocean and marine ecosystems. The movement of phosphate from the ocean to the land and through the soil is extremely slow, with the average phosphate ion having an oceanic residence time between 20,000 and 100,000 years.



In nature, phosphorus exists as the phosphate ion (PO_4^{3-}). Weathering of rocks and volcanic activity releases phosphate into the soil, water, and air, where it becomes available to terrestrial food webs. Phosphate enters the oceans via surface runoff, groundwater flow, and river flow. Phosphate dissolved in ocean water cycles into marine food webs. Some phosphate from the marine food webs falls to the ocean floor, where it

forms sediment. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

As discussed in [Chapter 44](#), excess phosphorus and nitrogen that enters these ecosystems from fertilizer runoff and from sewage causes excessive growth of microorganisms and depletes the dissolved oxygen, which leads to the death of many ecosystem fauna, such as shellfish and finfish. This process is responsible for dead zones in lakes and at the mouths of many major rivers ([link](#)).



Dead zones occur when phosphorus and nitrogen from fertilizers cause excessive growth of microorganisms, which depletes oxygen and kills fauna. Worldwide, large dead zones are found in coastal areas of high population density. (credit: NASA Earth Observatory)

As discussed earlier, a **dead zone** is an area within a freshwater or marine ecosystem where large areas are depleted of their normal flora and fauna; these zones can be caused by eutrophication, oil spills, dumping of toxic chemicals, and other human activities. The number of dead zones has been increasing for several years, and more than 400 of these zones were present as of 2008. One of the worst dead zones is off the coast of the United States in the Gulf of Mexico, where fertilizer runoff from the Mississippi River basin has created a dead zone of over 8463 square miles. Phosphate and nitrate runoff from fertilizers also negatively affect several lake and bay ecosystems including the Chesapeake Bay in the eastern United States.

Note:

Everyday Connection
Chesapeake Bay



(a)



(b)

This (a) satellite image shows the Chesapeake Bay, an ecosystem affected by phosphate and nitrate runoff. A (b) member of the Army Corps of Engineers holds a clump of oysters being used as a part of the oyster restoration effort in

the bay. (credit a: modification
of work by NASA/MODIS;
credit b: modification of work
by U.S. Army)

The Chesapeake Bay has long been valued as one of the most scenic areas on Earth; it is now in distress and is recognized as a declining ecosystem. In the 1970s, the Chesapeake Bay was one of the first ecosystems to have identified dead zones, which continue to kill many fish and bottom-dwelling species, such as clams, oysters, and worms. Several species have declined in the Chesapeake Bay due to surface water runoff containing excess nutrients from artificial fertilizer used on land. The source of the fertilizers (with high nitrogen and phosphate content) is not limited to agricultural practices. There are many nearby urban areas and more than 150 rivers and streams empty into the bay that are carrying fertilizer runoff from lawns and gardens. Thus, the decline of the Chesapeake Bay is a complex issue and requires the cooperation of industry, agriculture, and everyday homeowners.

Of particular interest to conservationists is the oyster population; it is estimated that more than 200,000 acres of oyster reefs existed in the bay in the 1700s, but that number has now declined to only 36,000 acres. Oyster harvesting was once a major industry for Chesapeake Bay, but it declined 88 percent between 1982 and 2007. This decline was due not only to fertilizer runoff and dead zones but also to overharvesting. Oysters require a certain minimum population density because they must be in close proximity to reproduce. Human activity has altered the oyster population and locations, greatly disrupting the ecosystem.

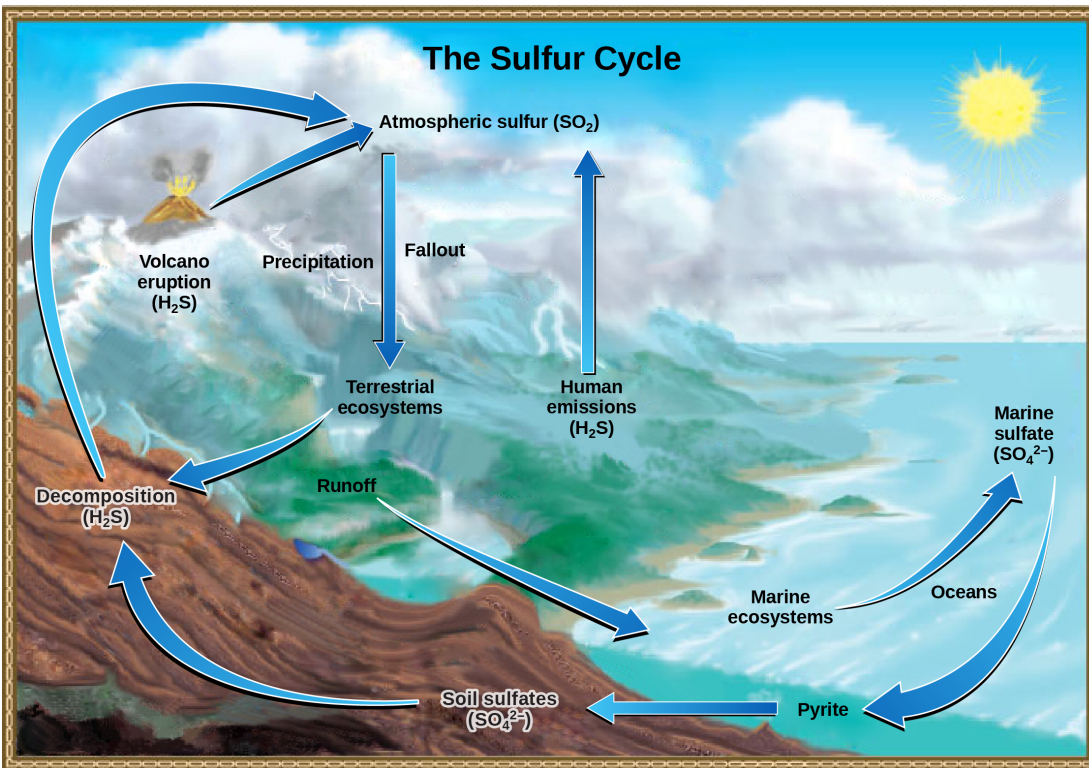
The restoration of the oyster population in the Chesapeake Bay has been ongoing for several years with mixed success. Not only do many people find oysters good to eat, but they also clean up the bay. Oysters are filter feeders, and as they eat, they clean the water around them. In the 1700s, it was estimated that it took only a few days for the oyster population to filter the entire volume of the bay. Today, with changed water conditions, it is estimated that the present population would take nearly a year to do the same job.

Restoration efforts have been ongoing for several years by nonprofit organizations, such as the Chesapeake Bay Foundation. The restoration goal is to find a way to increase population density so the oysters can reproduce more efficiently. Many disease-resistant varieties (developed at the Virginia Institute of Marine Science for the College of William and Mary) are now available and have been used in the construction of experimental oyster reefs. Efforts to clean and restore the bay by Virginia and Delaware have been hampered because much of the pollution entering the bay comes from other states, which stresses the need for interstate cooperation to gain successful restoration.

The new, hearty oyster strains have also spawned a new and economically viable industry—oyster aquaculture—which not only supplies oysters for food and profit, but also has the added benefit of cleaning the bay.

The Sulfur Cycle

Sulfur is an essential element for the macromolecules of living things. As a part of the amino acid cysteine, it is involved in the formation of disulfide bonds within proteins, which help to determine their 3-D folding patterns, and hence their functions. As shown in [\[link\]](#), sulfur cycles between the oceans, land, and atmosphere. Atmospheric sulfur is found in the form of sulfur dioxide (SO_2) and enters the atmosphere in three ways: from the decomposition of organic molecules, from volcanic activity and geothermal vents, and from the burning of fossil fuels by humans.



Sulfur dioxide from the atmosphere becomes available to terrestrial and marine ecosystems when it is dissolved in precipitation as weak sulfuric acid or when it falls directly to the Earth as fallout. Weathering of rocks also makes sulfates available to terrestrial ecosystems. Decomposition of living organisms returns sulfates to the ocean, soil, and atmosphere. (credit: modification of work by John M. Evans and Howard Perlman, USGS)

On land, sulfur is deposited in four major ways: precipitation, direct fallout from the atmosphere, rock weathering, and geothermal vents ([\[link\]](#)). Atmospheric sulfur is found in the form of sulfur dioxide (SO_2), and as rain falls through the atmosphere, sulfur is dissolved in the form of weak sulfuric acid (H_2SO_4). Sulfur can also fall directly from the atmosphere in a process called **fallout**. Also, the weathering of sulfur-containing rocks releases sulfur into the soil. These rocks originate from ocean sediments that are moved to land by the geologic uplifting of ocean sediments.

Terrestrial ecosystems can then make use of these soil sulfates (SO_4^-), and upon the death and decomposition of these organisms, release the sulfur back into the atmosphere as hydrogen sulfide (H_2S) gas.



At this sulfur vent in Lassen Volcanic National Park in northeastern California, the yellowish sulfur deposits are visible near the mouth of the vent.

Sulfur enters the ocean via runoff from land, from atmospheric fallout, and from underwater geothermal vents. Some ecosystems ([link](#)) rely on chemoautotrophs using sulfur as a biological energy source. This sulfur then supports marine ecosystems in the form of sulfates.

Human activities have played a major role in altering the balance of the global sulfur cycle. The burning of large quantities of fossil fuels, especially from coal, releases larger amounts of hydrogen sulfide gas into the atmosphere. **Acid rain** is caused by rainwater falling to the ground through this sulfur dioxide gas, turning it into weak sulfuric acid. Acid rain damages the natural environment by lowering the pH of lakes, which kills many of

the resident fauna; it also affects the man-made environment through the chemical degradation of buildings. For example, many marble monuments, such as the Lincoln Memorial in Washington, DC, have suffered significant damage from acid rain over the years.

Note:

Link to Learning

Click this [link](#) to learn more about global climate change.

Section Summary

Mineral nutrients are cycled through ecosystems and their environment. Of particular importance are water, carbon, nitrogen, phosphorus, and sulfur. All of these cycles have major impacts on ecosystem structure and function. A variety of human activities, such as pollution, oil spills, and other events have damaged ecosystems, potentially causing global climate change. The health of Earth depends on understanding these cycles and how to protect the environment from irreversible damage.

Visual Connection Questions

Exercise:

Problem:

[\[link\]](#) Which of the following statements about the nitrogen cycle is false?

- a. Ammonification converts organic nitrogenous matter from living organisms into ammonium (NH_4^+).
- b. Denitrification by bacteria converts nitrates (NO_3^-) to nitrogen gas (N_2).
- c. Nitrification by bacteria converts nitrates (NO_3^-) to nitrites (NO_2^-).

- d. Nitrogen fixing bacteria convert nitrogen gas (N_2) into organic compounds.
-

Solution:

[\[link\]](#) C: Nitrification by bacteria converts nitrates (NO_3^-) to nitrites (NO_2^-).

Review Questions

Exercise:

Problem:

The movement of mineral nutrients through organisms and their environment is called a _____ cycle.

- a. biological
- b. bioaccumulation
- c. biogeochemical
- d. biochemical

Solution:

C

Exercise:

Problem: Carbon is present in the atmosphere as _____.

- a. carbon dioxide
- b. carbonate ion
- c. carbon dust
- d. carbon monoxide

Solution:

A

Exercise:

Problem: The majority of water found on Earth is:

- a. ice
- b. water vapor
- c. fresh water
- d. salt water

Solution:

D

Exercise:

Problem:

The average time a molecule spends in its reservoir is known as _____.

- a. residence time
- b. restriction time
- c. resilience time
- d. storage time

Solution:

A

Exercise:

Problem:

The process whereby oxygen is depleted by the growth of microorganisms due to excess nutrients in aquatic systems is called _____.

- a. dead zoning
- b. eutrophication
- c. retrofication
- d. depletion

Solution:

B

Exercise:**Problem:**

The process whereby nitrogen is brought into organic molecules is called _____.

- a. nitrification
- b. denitrification
- c. nitrogen fixation
- d. nitrogen cycling

Solution:

C

Exercise:**Problem:**

Which of the following approaches would be the most effective way to reduce greenhouse carbon dioxide?

- a. Increase waste deposition into the deep ocean.
- b. Plant more environmentally-suitable plants.
- c. Increase use of fuel sources that do not produce carbon dioxide as a by-product.
- d. Decrease livestock agriculture.

Solution:

B

Exercise:

Problem:

How would loss of fungi in a forest effect biogeochemical cycles in the area?

- a. Nitrogen could no longer be fixed into organic molecules.
- b. Phosphorus stores would be released for use by other organisms.
- c. Sulfur release from eroding rocks would cease.
- d. Carbon would accumulate in dead organic matter and waste.

Solution:

D

Critical Thinking Questions

Exercise:

Problem:

Describe nitrogen fixation and why it is important to agriculture.

Solution:

Nitrogen fixation is the process of bringing nitrogen gas from the atmosphere and incorporating it into organic molecules. Most plants do not have this capability and must rely on free-living or symbiotic bacteria to do this. As nitrogen is often the limiting nutrient in the growth of crops, farmers make use of artificial fertilizers to provide a nitrogen source to the plants as they grow.

Exercise:

Problem:

What are the factors that cause dead zones? Describe eutrophication, in particular, as a cause.

Solution:

Many factors can kill life in a lake or ocean, such as eutrophication by nutrient-rich surface runoff, oil spills, toxic waste spills, changes in climate, and the dumping of garbage into the ocean. Eutrophication is a result of nutrient-rich runoff from land using artificial fertilizers high in nitrogen and phosphorus. These nutrients cause the rapid and excessive growth of microorganisms, which deplete local dissolved oxygen and kill many fish and other aquatic organisms.

Exercise:

Problem:

Why are drinking water supplies still a major concern for many countries?

Solution:

Most of the water on Earth is salt water, which humans cannot drink unless the salt is removed. Some fresh water is locked in glaciers and polar ice caps, or is present in the atmosphere. The Earth's water supplies are threatened by pollution and exhaustion. The effort to supply fresh drinking water to the planet's ever-expanding human population is seen as a major challenge in this century.

Exercise:**Problem:**

Discuss how the human disruption of the carbon cycle has caused ocean acidification.

Solution:

Human activity has greatly increased the amount of carbon dioxide gas in the Earth's atmosphere. The oceanic and atmospheric levels of carbon dioxide are linked so that when atmospheric carbon dioxide levels increase, the amount of dissolved carbon dioxide in the ocean also increases (partial pressure of oxygen). When carbon dioxide dissolves in water it produces the weak acid bicarbonate. Since the Industrial Revolution the pH of the ocean has dropped 0.1 pH units, a 30% increase in acidity.

Glossary**acid rain**

corrosive rain caused by rainwater falling to the ground through sulfur dioxide gas, turning it into weak sulfuric acid; can damage structures and ecosystems

biogeochemical cycle

cycling of mineral nutrients through ecosystems and through the nonliving world

dead zone

area within an ecosystem in lakes and near the mouths of rivers where large areas of ecosystems are depleted of their normal flora and fauna; these zones can be caused by eutrophication, oil spills, dumping of toxic chemicals, and other human activities

eutrophication

process whereby nutrient runoff causes the excess growth of microorganisms, depleting dissolved oxygen levels and killing

ecosystem fauna

fallout

direct deposit of solid minerals on land or in the ocean from the atmosphere

hydrosphere

area of the Earth where water movement and storage occurs

nonrenewable resource

resource, such as fossil fuel, that is either regenerated very slowly or not at all

residence time

measure of the average time an individual water molecule stays in a particular reservoir

subduction

movement of one tectonic plate beneath another

Energy Basics

By the end of this section, you will be able to:

- Define energy, distinguish types of energy, and describe the nature of energy changes that accompany chemical and physical changes
- Distinguish the related properties of heat, thermal energy, and temperature
- Define and distinguish specific heat and heat capacity, and describe the physical implications of both
- Perform calculations involving heat, specific heat, and temperature change

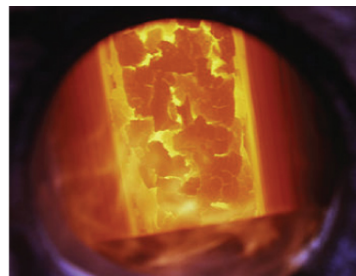
Chemical changes and their accompanying changes in energy are important parts of our everyday world ([\[link\]](#)). The macronutrients in food (proteins, fats, and carbohydrates) undergo metabolic reactions that provide the energy to keep our bodies functioning. We burn a variety of fuels (gasoline, natural gas, coal) to produce energy for transportation, heating, and the generation of electricity. Industrial chemical reactions use enormous amounts of energy to produce raw materials (such as iron and aluminum). Energy is then used to manufacture those raw materials into useful products, such as cars, skyscrapers, and bridges.



(a)



(b)



(c)

The energy involved in chemical changes is important to our daily lives: (a) A cheeseburger for lunch provides the energy you need to get through the rest of the day; (b) the combustion of gasoline provides the energy that moves your car (and you) between home, work, and school; and (c) coke, a processed form of coal, provides the energy needed to convert iron ore into iron, which is essential for making many of the products we use daily. (credit a: modification of work by

“Pink Sherbet Photography”/Flickr; credit b: modification of work by Jeffery Turner)

Over 90% of the energy we use comes originally from the sun. Every day, the sun provides the earth with almost 10,000 times the amount of energy necessary to meet all of the world’s energy needs for that day. Our challenge is to find ways to convert and store incoming solar energy so that it can be used in reactions or chemical processes that are both convenient and nonpolluting. Plants and many bacteria capture solar energy through photosynthesis. We release the energy stored in plants when we burn wood or plant products such as ethanol. We also use this energy to fuel our bodies by eating food that comes directly from plants or from animals that got their energy by eating plants. Burning coal and petroleum also releases stored solar energy: These fuels are fossilized plant and animal matter.

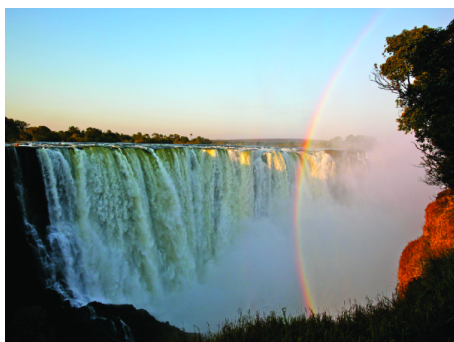
This chapter will introduce the basic ideas of an important area of science concerned with the amount of heat absorbed or released during chemical and physical changes—an area called **thermochemistry**. The concepts introduced in this chapter are widely used in almost all scientific and technical fields. Food scientists use them to determine the energy content of foods. Biologists study the energetics of living organisms, such as the metabolic combustion of sugar into carbon dioxide and water. The oil, gas, and transportation industries, renewable energy providers, and many others endeavor to find better methods to produce energy for our commercial and personal needs. Engineers strive to improve energy efficiency, find better ways to heat and cool our homes, refrigerate our food and drinks, and meet the energy and cooling needs of computers and electronics, among other applications. Understanding thermochemical principles is essential for chemists, physicists, biologists, geologists, every type of engineer, and just about anyone who studies or does any kind of science.

Energy

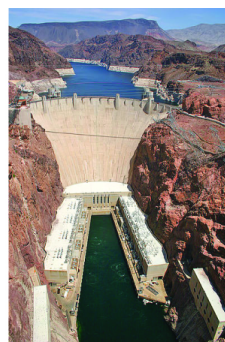
Energy can be defined as the capacity to supply heat or do work. One type of **work (w)** is the process of causing matter to move against an opposing force. For example, we do work when we inflate a bicycle tire—we move

matter (the air in the pump) against the opposing force of the air already in the tire.

Like matter, energy comes in different types. One scheme classifies energy into two types: **potential energy**, the energy an object has because of its relative position, composition, or condition, and **kinetic energy**, the energy that an object possesses because of its motion. Water at the top of a waterfall or dam has potential energy because of its position; when it flows downward through generators, it has kinetic energy that can be used to do work and produce electricity in a hydroelectric plant ([link](#)). A battery has potential energy because the chemicals within it can produce electricity that can do work.



(a)



(b)

(a) Water at a higher elevation, for example, at the top of Victoria Falls, has a higher potential energy than water at a lower elevation. As the water falls, some of its potential energy is converted into kinetic energy. (b) If the water flows through generators at the bottom of a dam, such as the Hoover Dam shown here, its kinetic energy is converted into electrical energy. (credit a: modification of work by Steve Jurvetson; credit b: modification of work by “curimedia”/Wikimedia commons)

Energy can be converted from one form into another, but all of the energy present before a change occurs always exists in some form after the change is completed. This observation is expressed in the law of conservation of energy: during a chemical or physical change, energy can be neither created nor destroyed, although it can be changed in form. (This is also one version of the first law of thermodynamics, as you will learn later.)

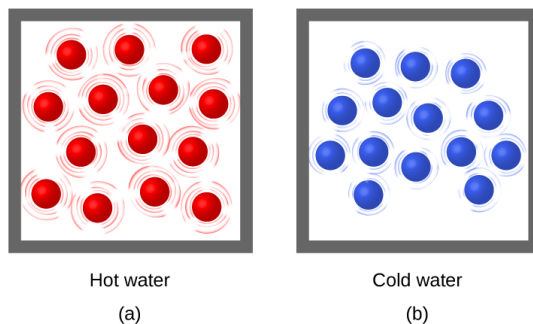
When one substance is converted into another, there is always an associated conversion of one form of energy into another. Heat is usually released or absorbed, but sometimes the conversion involves light, electrical energy, or some other form of energy. For example, chemical energy (a type of potential energy) is stored in the molecules that compose gasoline. When gasoline is combusted within the cylinders of a car's engine, the rapidly expanding gaseous products of this chemical reaction generate mechanical energy (a type of kinetic energy) when they move the cylinders' pistons.

According to the law of conservation of matter (seen in an earlier chapter), there is no detectable change in the total amount of matter during a chemical change. When chemical reactions occur, the energy changes are relatively modest and the mass changes are too small to measure, so the laws of conservation of matter and energy hold well. However, in nuclear reactions, the energy changes are much larger (by factors of a million or so), the mass changes are measurable, and matter-energy conversions are significant. This will be examined in more detail in a later chapter on nuclear chemistry.

Thermal Energy, Temperature, and Heat

Thermal energy is kinetic energy associated with the random motion of atoms and molecules. **Temperature** is a quantitative measure of “hot” or “cold.” When the atoms and molecules in an object are moving or vibrating quickly, they have a higher average kinetic energy (KE), and we say that the object is “hot.” When the atoms and molecules are moving slowly, they have lower average KE, and we say that the object is “cold” ([\[link\]](#)). Assuming that no chemical reaction or phase change (such as melting or vaporizing) occurs, increasing the amount of thermal energy in a sample of matter will cause its temperature to increase. And, assuming that no

chemical reaction or phase change (such as condensation or freezing) occurs, decreasing the amount of thermal energy in a sample of matter will cause its temperature to decrease.



(a) The molecules in a sample of hot water move more rapidly than (b) those in a sample of cold water.

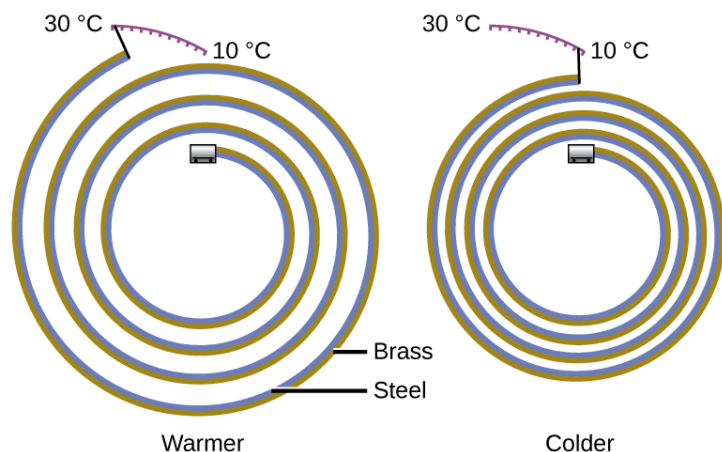
Note:

Click on this [interactive simulation](#) to view the effects of temperature on molecular motion.

Most substances expand as their temperature increases and contract as their temperature decreases. This property can be used to measure temperature changes, as shown in [\[link\]](#). The operation of many thermometers depends on the expansion and contraction of substances in response to temperature changes.



(a)



(b)

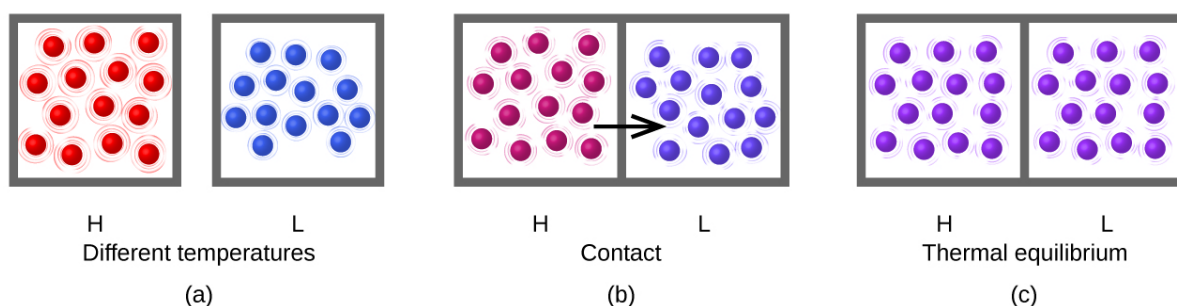
(a) In an alcohol or mercury thermometer, the liquid (dyed red for visibility) expands when heated and contracts when cooled, much more so than the glass tube that contains the liquid. (b) In a bimetallic thermometer, two different metals (such as brass and steel) form a two-layered strip. When heated or cooled, one of the metals (brass) expands or contracts more than the other metal (steel), causing the strip to coil or uncoil. Both types of thermometers have a calibrated scale that indicates the temperature. (credit a: modification of work by “dwstucke”/Flickr)

Note:

The following [demonstration](#) allows one to view the effects of heating and cooling a coiled bimetallic strip.

Heat (q) is the transfer of thermal energy between two bodies at different temperatures. Heat flow (a redundant term, but one commonly used) increases the thermal energy of one body and decreases the thermal energy of the other. Suppose we initially have a high temperature (and high thermal energy) substance (H) and a low temperature (and low thermal energy)

substance (L). The atoms and molecules in H have a higher average KE than those in L. If we place substance H in contact with substance L, the thermal energy will flow spontaneously from substance H to substance L. The temperature of substance H will decrease, as will the average KE of its molecules; the temperature of substance L will increase, along with the average KE of its molecules. Heat flow will continue until the two substances are at the same temperature ([link](#)).



(a) Substances H and L are initially at different temperatures, and their atoms have different average kinetic energies. (b) When they contact each other, collisions between the molecules result in the transfer of kinetic (thermal) energy from the hotter to the cooler matter. (c) The two objects reach “thermal equilibrium” when both substances are at the same temperature and their molecules have the same average kinetic energy.

Note:

Click on the [PhET simulation](#) to explore energy forms and changes. Visit the Energy Systems tab to create combinations of energy sources, transformation methods, and outputs. Click on Energy Symbols to visualize the transfer of energy.

Matter undergoing chemical reactions and physical changes can release or absorb heat. A change that releases heat is called an **exothermic process**. For example, the combustion reaction that occurs when using an oxyacetylene torch is an exothermic process—this process also releases energy in the form of light as evidenced by the torch's flame ([link](#)). A reaction or change that absorbs heat is an **endothermic process**. A cold pack used to treat muscle strains provides an example of an endothermic process. When the substances in the cold pack (water and a salt like ammonium nitrate) are brought together, the resulting process absorbs heat, leading to the sensation of cold.



(a)



(b)

(a) An oxyacetylene torch produces heat by the combustion of acetylene in oxygen. The energy released by this exothermic reaction heats and then melts the metal being cut. The sparks are tiny bits of the molten metal flying away. (b) A cold pack uses an endothermic process to create the sensation of cold. (credit a: modification of work by “Skatebiker”/Wikimedia commons)

Historically, energy was measured in units of **calories (cal)**. A calorie is the amount of energy required to raise one gram of water by 1 degree C (1 kelvin). However, this quantity depends on the atmospheric pressure and

the starting temperature of the water. The ease of measurement of energy changes in calories has meant that the calorie is still frequently used. The Calorie (with a capital C), or large calorie, commonly used in quantifying food energy content, is a kilocalorie. The SI unit of heat, work, and energy is the joule. A **joule (J)** is defined as the amount of energy used when a force of 1 newton moves an object 1 meter. It is named in honor of the English physicist James Prescott Joule. One joule is equivalent to $1 \text{ kg m}^2/\text{s}^2$, which is also called 1 newton–meter. A kilojoule (kJ) is 1000 joules. To standardize its definition, 1 calorie has been set to equal 4.184 joules.

We now introduce two concepts useful in describing heat flow and temperature change. The **heat capacity (C)** of a body of matter is the quantity of heat (q) it absorbs or releases when it experiences a temperature change (ΔT) of 1 degree Celsius (or equivalently, 1 kelvin):

Equation:

$$C = \frac{q}{\Delta T}$$

Heat capacity is determined by both the type and amount of substance that absorbs or releases heat. It is therefore an extensive property—its value is proportional to the amount of the substance.

For example, consider the heat capacities of two cast iron frying pans. The heat capacity of the large pan is five times greater than that of the small pan because, although both are made of the same material, the mass of the large pan is five times greater than the mass of the small pan. More mass means more atoms are present in the larger pan, so it takes more energy to make all of those atoms vibrate faster. The heat capacity of the small cast iron frying pan is found by observing that it takes 18,150 J of energy to raise the temperature of the pan by 50.0 °C:

Equation:

$$C_{\text{small pan}} = \frac{18,140 \text{ J}}{50.0 \text{ }^{\circ}\text{C}} = 363 \text{ J/}^{\circ}\text{C}$$

The larger cast iron frying pan, while made of the same substance, requires 90,700 J of energy to raise its temperature by 50.0 °C. The larger pan has a (proportionally) larger heat capacity because the larger amount of material requires a (proportionally) larger amount of energy to yield the same temperature change:

Equation:

$$C_{\text{large pan}} = \frac{90,700 \text{ J}}{50.0 \text{ }^{\circ}\text{C}} = 1814 \text{ J/}^{\circ}\text{C}$$

The **specific heat capacity (c)** of a substance, commonly called its “specific heat,” is the quantity of heat required to raise the temperature of 1 gram of a substance by 1 degree Celsius (or 1 kelvin):

Equation:

$$c = \frac{q}{m\Delta T}$$

Specific heat capacity depends only on the kind of substance absorbing or releasing heat. It is an intensive property—the type, but not the amount, of the substance is all that matters. For example, the small cast iron frying pan has a mass of 808 g. The specific heat of iron (the material used to make the pan) is therefore:

Equation:

$$c_{\text{iron}} = \frac{18,140 \text{ J}}{(808 \text{ g})(50.0 \text{ }^{\circ}\text{C})} = 0.449 \text{ J/g }^{\circ}\text{C}$$

The large frying pan has a mass of 4040 g. Using the data for this pan, we can also calculate the specific heat of iron:

Equation:

$$c_{\text{iron}} = \frac{90,700 \text{ J}}{(4040 \text{ g})(50.0 \text{ }^{\circ}\text{C})} = 0.449 \text{ J/g }^{\circ}\text{C}$$

Although the large pan is more massive than the small pan, since both are made of the same material, they both yield the same value for specific heat (for the material of construction, iron). Note that specific heat is measured in units of energy per temperature per mass and is an intensive property, being derived from a ratio of two extensive properties (heat and mass). The molar heat capacity, also an intensive property, is the heat capacity per mole of a particular substance and has units of $\text{J/mol } ^\circ\text{C}$ ([\[link\]](#)).



Because of its larger mass, a large frying pan has a larger heat capacity than a small frying pan. Because they are made of the same material, both frying pans have the same specific heat. (credit: Mark Blaser)

water has a relatively high specific heat (about $4.2 \text{ J/g } ^\circ\text{C}$ for the liquid and $2.09 \text{ J/g } ^\circ\text{C}$ for the solid)); most metals have much lower specific heats (usually less than $1 \text{ J/g } ^\circ\text{C}$). The specific heat of a substance varies somewhat with temperature. However, this variation is usually small enough that we will treat specific heat as constant over the range of temperatures that will be considered in this chapter. Specific heats of some common substances are listed in [\[link\]](#).

Specific Heats of Common Substances at 25 °C and 1 bar		
Substance	Symbol (<i>state</i>)	Specific Heat (J/g °C)
helium	He(<i>g</i>)	5.193
water	H ₂ O(<i>l</i>)	4.184
ethanol	C ₂ H ₆ O(<i>l</i>)	2.376
ice	H ₂ O(<i>s</i>)	2.093 (at −10 °C)
water vapor	H ₂ O(<i>g</i>)	1.864
nitrogen	N ₂ (<i>g</i>)	1.040
air		1.007
oxygen	O ₂ (<i>g</i>)	0.918
aluminum	Al(<i>s</i>)	0.897
carbon dioxide	CO ₂ (<i>g</i>)	0.853
argon	Ar(<i>g</i>)	0.522
iron	Fe(<i>s</i>)	0.449
copper	Cu(<i>s</i>)	0.385
lead	Pb(<i>s</i>)	0.130
gold	Au(<i>s</i>)	0.129
silicon	Si(<i>s</i>)	0.712

If we know the mass of a substance and its specific heat, we can determine the amount of heat, q , entering or leaving the substance by measuring the temperature change before and after the heat is gained or lost:

Equation:

$$q = (\text{specific heat}) \times (\text{mass of substance}) \times (\text{temperature change})$$
$$q = c \times m \times \Delta T = c \times m \times (T_{\text{final}} - T_{\text{initial}})$$

In this equation, c is the specific heat of the substance, m is its mass, and ΔT (which is read “delta T”) is the temperature change, $T_{\text{final}} - T_{\text{initial}}$. If a substance gains thermal energy, its temperature increases, its final temperature is higher than its initial temperature, $T_{\text{final}} - T_{\text{initial}}$ has a positive value, and the value of q is positive. If a substance loses thermal energy, its temperature decreases, the final temperature is lower than the initial temperature, $T_{\text{final}} - T_{\text{initial}}$ has a negative value, and the value of q is negative.

Example:

Measuring Heat

A flask containing 8.0×10^2 g of water is heated, and the temperature of the water increases from 21 °C to 85 °C. How much heat did the water absorb?

Solution

To answer this question, consider these factors:

- the specific heat of the substance being heated (in this case, water)
- the amount of substance being heated (in this case, 8.0×10^2 g)
- the magnitude of the temperature change (in this case, from 21 °C to 85 °C).

The specific heat of water is 4.184 J/g °C, so to heat 1 g of water by 1 °C requires 4.184 J. We note that since 4.184 J is required to heat 1 g of water by 1 °C, we will need *800 times as much* to heat 8.0×10^2 g of water by 1 °C. Finally, we observe that since 4.184 J are required to heat 1 g of water

by 1 °C, we will need *64 times as much* to heat it by 64 °C (that is, from 21 °C to 85 °C).

This can be summarized using the equation:

Equation:

$$q = c \times m \times \Delta T = c \times m \times (T_{\text{final}} - T_{\text{initial}})$$

Equation:

$$\begin{aligned} &= (4.184 \text{ J/g } ^\circ\text{C}) \times (8.0 \times 10^2 \text{ g}) \times (85 - 21) ^\circ\text{C} \\ &= (4.184 \text{ J/g } ^\circ\text{C}) \times (8.0 \times 10^2 \text{ g}) \times (64) ^\circ\text{C} \\ &= 210,000 \text{ J } (= 2.1 \times 10^5 \text{ kJ}) \end{aligned}$$

Because the temperature increased, the water absorbed heat and q is positive.

Check Your Learning

How much heat, in joules, must be added to a $5.07 \times 10^4 \text{ g}$ iron skillet to increase its temperature from 25 °C to 250 °C? The specific heat of iron is 0.449 J/g °C.

Note:

Answer:

$$5.07 \times 10^4 \text{ J}$$

Note that the relationship between heat, specific heat, mass, and temperature change can be used to determine any of these quantities (not just heat) if the other three are known or can be deduced.

Example:**Determining Other Quantities**

A piece of unknown metal weighs 348 g. When the metal piece absorbs 6.64 kJ of heat, its temperature increases from 22.4 °C to 43.6 °C.

Determine the specific heat of this metal (which might provide a clue to its identity).

Solution

Since mass, heat, and temperature change are known for this metal, we can determine its specific heat using the relationship:

Equation:

$$q = c \times m \times \Delta T = c \times m \times (T_{\text{final}} - T_{\text{initial}})$$

Substituting the known values:

Equation:

$$6640 \text{ J} = c \times (348 \text{ g}) \times (43.6 - 22.4) ^\circ\text{C}$$

Solving:

Equation:

$$c = \frac{6640 \text{ J}}{(348 \text{ g}) \times (21.2 ^\circ\text{C})} = 0.900 \text{ J/g } ^\circ\text{C}$$

Comparing this value with the values in [\[link\]](#), this value matches the specific heat of aluminum, which suggests that the unknown metal may be aluminum.

Check Your Learning

A piece of unknown metal weighs 217 g. When the metal piece absorbs 1.43 kJ of heat, its temperature increases from 24.5 °C to 39.1 °C.

Determine the specific heat of this metal, and predict its identity.

Note:**Answer:**

$c = 0.451 \text{ J/g } ^\circ\text{C}$; the metal is likely to be iron

Note:

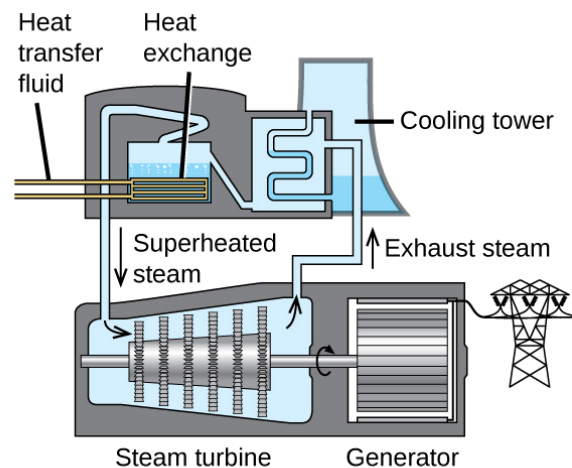
Solar Thermal Energy Power Plants

The sunlight that reaches the earth contains thousands of times more energy than we presently capture. Solar thermal systems provide one possible solution to the problem of converting energy from the sun into energy we can use. Large-scale solar thermal plants have different design specifics, but all concentrate sunlight to heat some substance; the heat “stored” in that substance is then converted into electricity.

The Solana Generating Station in Arizona’s Sonora Desert produces 280 megawatts of electrical power. It uses parabolic mirrors that focus sunlight on pipes filled with a heat transfer fluid (HTF) ([link](#)). The HTF then does two things: It turns water into steam, which spins turbines, which in turn produces electricity, and it melts and heats a mixture of salts, which functions as a thermal energy storage system. After the sun goes down, the molten salt mixture can then release enough of its stored heat to produce steam to run the turbines for 6 hours. Molten salts are used because they possess a number of beneficial properties, including high heat capacities and thermal conductivities.



(a)



(b)

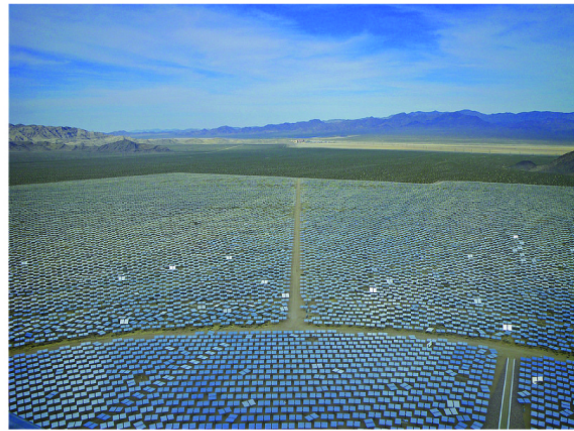
This solar thermal plant uses parabolic trough mirrors to concentrate sunlight. (credit a: modification of work by Bureau of Land

Management)

The 377-megawatt Ivanpah Solar Generating System, located in the Mojave Desert in California, is the largest solar thermal power plant in the world ([\[link\]](#)). Its 170,000 mirrors focus huge amounts of sunlight on three water-filled towers, producing steam at over 538 °C that drives electricity-producing turbines. It produces enough energy to power 140,000 homes. Water is used as the working fluid because of its large heat capacity and heat of vaporization.



(a)



(b)

(a) The Ivanpah solar thermal plant uses 170,000 mirrors to concentrate sunlight on water-filled towers. (b) It covers 4000 acres of public land near the Mojave Desert and the California-Nevada border.

(credit a: modification of work by Craig Dietrich; credit b: modification of work by “USFWS Pacific Southwest Region”/Flickr)

Key Concepts and Summary

Energy is the capacity to supply heat or do work (applying a force to move matter). Kinetic energy (KE) is the energy of motion; potential energy is energy due to relative position, composition, or condition. When energy is

converted from one form into another, energy is neither created nor destroyed (law of conservation of energy or first law of thermodynamics).

The thermal energy of matter is due to the kinetic energies of its constituent atoms or molecules. Temperature is an intensive property of matter reflecting hotness or coldness that increases as the average kinetic energy increases. Heat is the transfer of thermal energy between objects at different temperatures. Chemical and physical processes can absorb heat (endothermic) or release heat (exothermic). The SI unit of energy, heat, and work is the joule (J).

Specific heat and heat capacity are measures of the energy needed to change the temperature of a substance or object. The amount of heat absorbed or released by a substance depends directly on the type of substance, its mass, and the temperature change it undergoes.

Key Equations

- $q = c \times m \times \Delta T = c \times m \times (T_{\text{final}} - T_{\text{initial}})$

Chemistry End of Chapter Exercises

Exercise:

Problem:

A burning match and a bonfire may have the same temperature, yet you would not sit around a burning match on a fall evening to stay warm. Why not?

Solution:

The temperature of 1 gram of burning wood is approximately the same for both a match and a bonfire. This is an intensive property and depends on the material (wood). However, the overall amount of produced heat depends on the amount of material; this is an extensive property. The amount of wood in a bonfire is much greater than that in

a match; the total amount of produced heat is also much greater, which is why we can sit around a bonfire to stay warm, but a match would not provide enough heat to keep us from getting cold.

Exercise:

Problem:

Prepare a table identifying several energy transitions that take place during the typical operation of an automobile.

Exercise:

Problem:

Explain the difference between heat capacity and specific heat of a substance.

Solution:

Heat capacity refers to the heat required to raise the temperature of the mass of the substance 1 degree; specific heat refers to the heat required to raise the temperature of 1 gram of the substance 1 degree. Thus, heat capacity is an extensive property, and specific heat is an intensive one.

Exercise:

Problem:

Calculate the heat capacity, in joules and in calories per degree, of the following:

(a) 28.4 g of water

(b) 1.00 oz of lead

Exercise:

Problem:

Calculate the heat capacity, in joules and in calories per degree, of the following:

(a) 45.8 g of nitrogen gas

(b) 1.00 pound of aluminum metal

Solution:

(a) 47.6 J/°C; 11.38 cal °C⁻¹; (b) 407 J/°C; 97.3 cal °C⁻¹

Exercise:

Problem:

How much heat, in joules and in calories, must be added to a 75.0-g iron block with a specific heat of 0.449 J/g °C to increase its temperature from 25 °C to its melting temperature of 1535 °C?

Exercise:

Problem:

How much heat, in joules and in calories, is required to heat a 28.4-g (1-oz) ice cube from -23.0 °C to -1.0 °C?

Solution:

1310 J; 313 cal

Exercise:

Problem:

How much would the temperature of 275 g of water increase if 36.5 kJ of heat were added?

Exercise:

Problem:

If 14.5 kJ of heat were added to 485 g of liquid water, how much would its temperature increase?

Solution:

7.15 °C

Exercise:

Problem:

A piece of unknown substance weighs 44.7 g and requires 2110 J to increase its temperature from 23.2 °C to 89.6 °C.

- (a) What is the specific heat of the substance?
- (b) If it is one of the substances found in [\[link\]](#), what is its likely identity?

Exercise:

Problem:

A piece of unknown solid substance weighs 437.2 g, and requires 8460 J to increase its temperature from 19.3 °C to 68.9 °C.

- (a) What is the specific heat of the substance?
- (b) If it is one of the substances found in [\[link\]](#), what is its likely identity?

Solution:

- (a) 0.390 J/g °C; (b) Copper is a likely candidate.

Exercise:

Problem: An aluminum kettle weighs 1.05 kg.

- (a) What is the heat capacity of the kettle?
- (b) How much heat is required to increase the temperature of this kettle from 23.0 °C to 99.0 °C?
- (c) How much heat is required to heat this kettle from 23.0 °C to 99.0 °C if it contains 1.25 L of water (density of 0.997 g/mL and a specific

heat of 4.184 J/g °C)?

Exercise:

Problem:

Most people find waterbeds uncomfortable unless the water temperature is maintained at about 85 °F. Unless it is heated, a waterbed that contains 892 L of water cools from 85 °F to 72 °F in 24 hours. Estimate the amount of electrical energy required over 24 hours, in kWh, to keep the bed from cooling. Note that 1 kilowatt-hour (kWh) = 3.6×10^6 J, and assume that the density of water is 1.0 g/mL (independent of temperature). What other assumptions did you make? How did they affect your calculated result (i.e., were they likely to yield “positive” or “negative” errors)?

Solution:

We assume that the density of water is 1.0 g/cm³(1 g/mL) and that it takes as much energy to keep the water at 85 °F as to heat it from 72 °F to 85 °F. We also assume that only the water is going to be heated.
Energy required = 7.47 kWh

Glossary

calorie (cal)

unit of heat or other energy; the amount of energy required to raise 1 gram of water by 1 degree Celsius; 1 cal is defined as 4.184 J

endothermic process

chemical reaction or physical change that absorbs heat

energy

capacity to supply heat or do work

exothermic process

chemical reaction or physical change that releases heat

heat (q)

transfer of thermal energy between two bodies

heat capacity (C)

extensive property of a body of matter that represents the quantity of heat required to increase its temperature by 1 degree Celsius (or 1 kelvin)

joule (J)

SI unit of energy; 1 joule is the kinetic energy of an object with a mass of 2 kilograms moving with a velocity of 1 meter per second, $1 \text{ J} = 1 \text{ kg m}^2/\text{s}^2$ and $4.184 \text{ J} = 1 \text{ cal}$

kinetic energy

energy of a moving body, in joules, equal to $\frac{1}{2}mv^2$ (where m = mass and v = velocity)

potential energy

energy of a particle or system of particles derived from relative position, composition, or condition

specific heat capacity (c)

intensive property of a substance that represents the quantity of heat required to raise the temperature of 1 gram of the substance by 1 degree Celsius (or 1 kelvin)

temperature

intensive property of matter that is a quantitative measure of “hotness” and “coldness”

thermal energy

kinetic energy associated with the random motion of atoms and molecules

thermochemistry

study of measuring the amount of heat absorbed or released during a chemical reaction or a physical change

work (w)

energy transfer due to changes in external, macroscopic variables such as pressure and volume; or causing matter to move against an opposing force

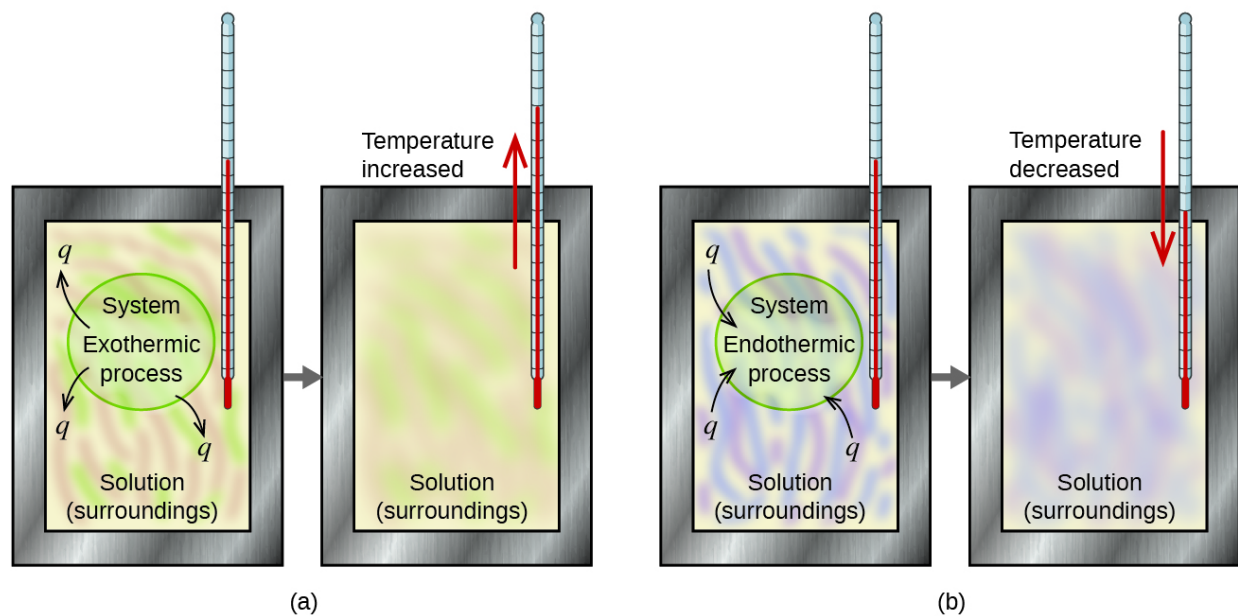
Calorimetry

By the end of this section, you will be able to:

- Explain the technique of calorimetry
- Calculate and interpret heat and related properties using typical calorimetry data

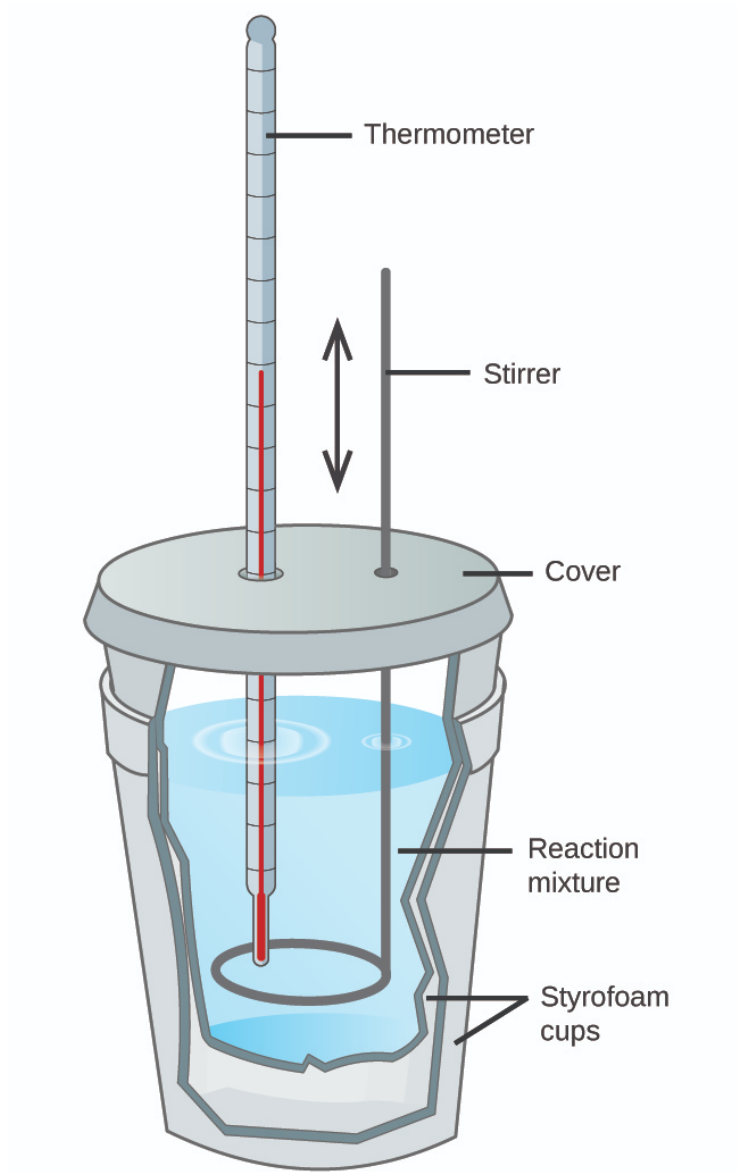
One technique we can use to measure the amount of heat involved in a chemical or physical process is known as **calorimetry**. Calorimetry is used to measure amounts of heat transferred to or from a substance. To do so, the heat is exchanged with a calibrated object (calorimeter). The temperature change measured by the calorimeter is used to derive the amount of heat transferred by the process under study. The measurement of heat transfer using this approach requires the definition of a **system** (the substance or substances undergoing the chemical or physical change) and its **surroundings** (all other matter, including components of the measurement apparatus, that serve to either provide heat to the system or absorb heat from the system).

A **calorimeter** is a device used to measure the amount of heat involved in a chemical or physical process. For example, when an exothermic reaction occurs in solution in a calorimeter, the heat produced by the reaction is absorbed by the solution, which increases its temperature. When an endothermic reaction occurs, the heat required is absorbed from the thermal energy of the solution, which decreases its temperature ([link](#)). The temperature change, along with the specific heat and mass of the solution, can then be used to calculate the amount of heat involved in either case.



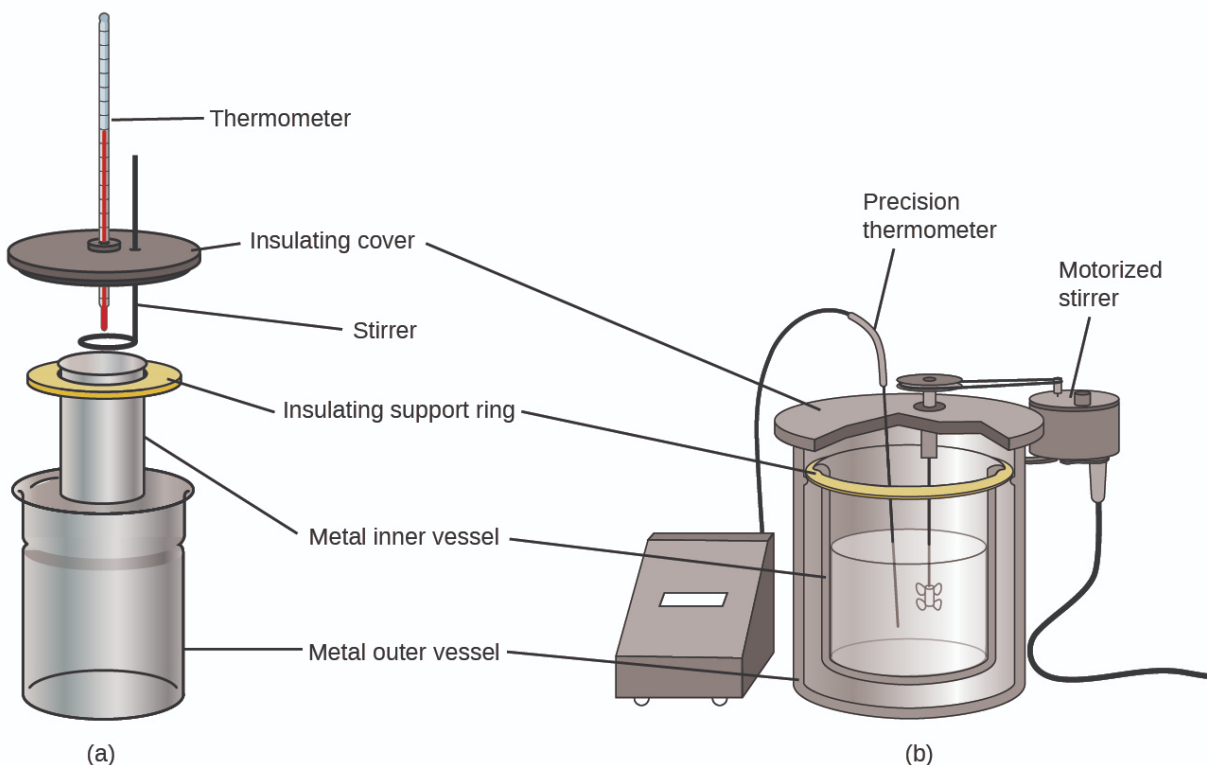
In a calorimetric determination, either (a) an exothermic process occurs and heat, q , is negative, indicating that thermal energy is transferred from the system to its surroundings, or (b) an endothermic process occurs and heat, q , is positive, indicating that thermal energy is transferred from the surroundings to the system.

Scientists use well-insulated calorimeters that all but prevent the transfer of heat between the calorimeter and its environment, which effectively limits the “surroundings” to the nonsystem components with the calorimeter (and the calorimeter itself). This enables the accurate determination of the heat involved in chemical processes, the energy content of foods, and so on. General chemistry students often use simple calorimeters constructed from polystyrene cups ([link](#)). These easy-to-use “coffee cup” calorimeters allow more heat exchange with the outside environment, and therefore produce less accurate energy values.



A simple calorimeter can be constructed from two polystyrene cups. A thermometer and stirrer extend through the cover into the reaction mixture.

Commercial solution calorimeters are also available. Relatively inexpensive calorimeters often consist of two thin-walled cups that are nested in a way that minimizes thermal contact during use, along with an insulated cover, handheld stirrer, and simple thermometer. More expensive calorimeters used for industry and research typically have a well-insulated, fully enclosed reaction vessel, motorized stirring mechanism, and a more accurate temperature sensor ([link](#)).



Commercial solution calorimeters range from (a) simple, inexpensive models for student use to (b) expensive, more accurate models for industry and research.

Before discussing the calorimetry of chemical reactions, consider a simpler example that illustrates the core idea behind calorimetry. Suppose we initially have a high-temperature substance, such as a hot piece of metal (M), and a low-temperature substance, such as cool water (W). If we place the metal in the water, heat will flow from M to W. The temperature of M will decrease, and the temperature of W will increase, until the two substances have the same temperature—that is, when they reach thermal equilibrium ([link](#)). If this occurs in a calorimeter, ideally all of this heat transfer occurs between the two substances, with no heat gained or lost by either its external environment. Under these ideal circumstances, the net heat change is zero:

Equation:

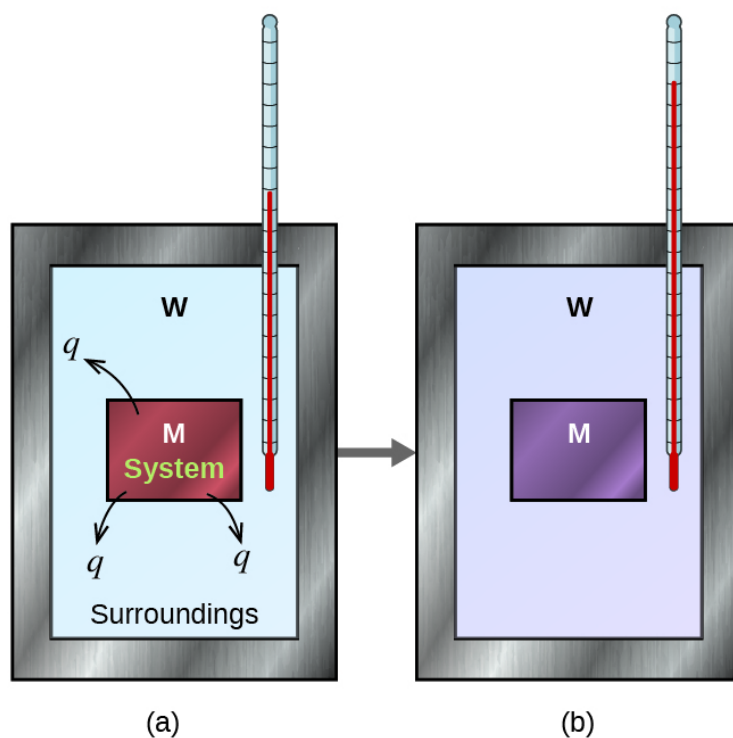
$$q_{\text{substance M}} + q_{\text{substance W}} = 0$$

This relationship can be rearranged to show that the heat gained by substance M is equal to the heat lost by substance W:

Equation:

$$q_{\text{substance M}} = -q_{\text{substance W}}$$

The magnitude of the heat (change) is therefore the same for both substances, and the negative sign merely shows that $q_{\text{substance M}}$ and $q_{\text{substance W}}$ are opposite in direction of heat flow (gain or loss) but does not indicate the arithmetic sign of either q value (that is determined by whether the matter in question gains or loses heat, per definition). In the specific situation described, $q_{\text{substance M}}$ is a negative value and $q_{\text{substance W}}$ is positive, since heat is transferred from M to W.



In a simple calorimetry process, (a) heat, q , is transferred from the hot metal, M, to the cool water, W, until (b) both are at the same temperature.

Example:

Heat Transfer between Substances at Different Temperatures

A 360.0-g piece of rebar (a steel rod used for reinforcing concrete) is dropped into 425 mL of water at 24.0 °C. The final temperature of the water was measured as 42.7 °C. Calculate the initial temperature of the piece of rebar. Assume the specific heat of steel is approximately the same as that for iron ([link](#)), and that all heat transfer occurs between the rebar and the water (there is no heat exchange with the surroundings).

Solution

The temperature of the water increases from 24.0 °C to 42.7 °C, so the water absorbs heat. That heat came from the piece of rebar, which initially was at a higher temperature. Assuming that all heat transfer was between the rebar and the water, with no heat “lost” to the outside environment, then *heat given off by rebar* = *–heat taken in by water*, or:

Equation:

$$q_{\text{rebar}} = -q_{\text{water}}$$

Since we know how heat is related to other measurable quantities, we have:

Equation:

$$(c \times m \times \Delta T)_{\text{rebar}} = -(c \times m \times \Delta T)_{\text{water}}$$

Letting f = final and i = initial, in expanded form, this becomes:

Equation:

$$c_{\text{rebar}} \times m_{\text{rebar}} \times (T_{\text{f, rebar}} - T_{\text{i, rebar}}) = -c_{\text{water}} \times m_{\text{water}} \times (T_{\text{f, water}} - T_{\text{i, water}})$$

The density of water is 1.0 g/mL, so 425 mL of water = 425 g. Noting that the final temperature of both the rebar and water is 42.7 °C, substituting known values yields:

Equation:

$$(0.449 \text{ J/g } ^\circ\text{C})(360.0 \text{ g})(42.7 ^\circ\text{C} - T_{\text{i, rebar}}) = -(4.184 \text{ J/g } ^\circ\text{C})(425 \text{ g})(42.7 ^\circ\text{C} - 24.0 ^\circ\text{C})$$

Equation:

$$T_{\text{i, rebar}} = \frac{(4.184 \text{ J/g } ^\circ\text{C})(425 \text{ g})(42.7 ^\circ\text{C} - 24.0 ^\circ\text{C})}{(0.449 \text{ J/g } ^\circ\text{C})(360.0 \text{ g})} + 42.7 ^\circ\text{C}$$

Solving this gives $T_{\text{i, rebar}} = 248 ^\circ\text{C}$, so the initial temperature of the rebar was 248 °C.

Check Your Learning

A 248-g piece of copper is dropped into 390 mL of water at 22.6 °C. The final temperature of the water was measured as 39.9 °C. Calculate the initial temperature of the piece of copper. Assume that all heat transfer occurs between the copper and the water.

Note:

Answer:

The initial temperature of the copper was 335.6 °C.

Check Your Learning

A 248-g piece of copper initially at 314 °C is dropped into 390 mL of water initially at 22.6 °C. Assuming that all heat transfer occurs between the copper and the water, calculate the final temperature.

Note:

Answer:

The final temperature (reached by both copper and water) is 38.7 °C.

This method can also be used to determine other quantities, such as the specific heat of an unknown metal.

Example:

Identifying a Metal by Measuring Specific Heat

A 59.7 g piece of metal that had been submerged in boiling water was quickly transferred into 60.0 mL of water initially at 22.0 °C. The final temperature is 28.5 °C. Use these data to determine the specific heat of the metal. Use this result to identify the metal.

Solution

Assuming perfect heat transfer, *heat given off by metal* = *−heat taken in by water*, or:

Equation:

$$q_{\text{metal}} = -q_{\text{water}}$$

In expanded form, this is:

Equation:

$$c_{\text{metal}} \times m_{\text{metal}} \times (T_{\text{f,metal}} - T_{\text{i, metal}}) = -c_{\text{water}} \times m_{\text{water}} \times (T_{\text{f,water}} - T_{\text{i,water}})$$

Noting that since the metal was submerged in boiling water, its initial temperature was 100.0 °C; and that for water, 60.0 mL = 60.0 g; we have:

Equation:

$$(c_{\text{metal}})(59.7 \text{ g})(28.5 \text{ }^{\circ}\text{C} - 100.0 \text{ }^{\circ}\text{C}) = -(4.18 \text{ J/g }^{\circ}\text{C})(60.0 \text{ g})(28.5 \text{ }^{\circ}\text{C} - 22.0 \text{ }^{\circ}\text{C})$$

Solving this:

Equation:

$$c_{\text{metal}} = \frac{-(4.184 \text{ J/g }^{\circ}\text{C})(60.0 \text{ g})(6.5 \text{ }^{\circ}\text{C})}{(59.7 \text{ g})(-71.5 \text{ }^{\circ}\text{C})} = 0.38 \text{ J/g }^{\circ}\text{C}$$

Comparing this with values in [\[link\]](#), our experimental specific heat is closest to the value for copper (0.39 J/g °C), so we identify the metal as copper.

Check Your Learning

A 92.9-g piece of a silver/gray metal is heated to 178.0 °C, and then quickly transferred into 75.0 mL of water initially at 24.0 °C. After 5 minutes, both the metal and the water have reached the same temperature: 29.7 °C. Determine the specific heat and the identity of the metal. (Note: You should find that the specific heat is close to that of two different metals. Explain how you can confidently determine the identity of the metal).

Note:

Answer:

$$c_{\text{metal}} = 0.13 \text{ J/g }^{\circ}\text{C}$$

This specific heat is close to that of either gold or lead. It would be difficult to determine which metal this was based solely on the numerical values. However, the observation that the metal is silver/gray in addition to the value for the specific heat indicates that the metal is lead.

When we use calorimetry to determine the heat involved in a chemical reaction, the same principles we have been discussing apply. The amount of heat absorbed by the calorimeter is often small enough that we can neglect it (though not for highly accurate measurements, as discussed later), and the calorimeter minimizes energy exchange with the outside environment. Because energy is neither created nor destroyed during a chemical reaction, the heat produced or consumed in the reaction (the “system”), q_{reaction} , plus the heat absorbed or lost by the solution (the “surroundings”), q_{solution} , must add up to zero:

Equation:

$$q_{\text{reaction}} + q_{\text{solution}} = 0$$

This means that the amount of heat produced or consumed in the reaction equals the amount of heat absorbed or lost by the solution:

Equation:

$$q_{\text{reaction}} = -q_{\text{solution}}$$

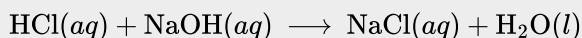
This concept lies at the heart of all calorimetry problems and calculations.

Example:

Heat Produced by an Exothermic Reaction

When 50.0 mL of 1.00 M HCl(aq) and 50.0 mL of 1.00 M NaOH(aq), both at 22.0 °C, are added to a coffee cup calorimeter, the temperature of the mixture reaches a maximum of 28.9 °C. What is the approximate amount of heat produced by this reaction?

Equation:



Solution

To visualize what is going on, imagine that you could combine the two solutions so quickly that no reaction took place while they mixed; then after mixing, the reaction took place. At the instant of mixing, you have 100.0 mL of a mixture of HCl and NaOH at 22.0 °C. The HCl and NaOH then react until the solution temperature reaches 28.9 °C.

The heat given off by the reaction is equal to that taken in by the solution. Therefore:

Equation:

$$q_{\text{reaction}} = -q_{\text{solution}}$$

(It is important to remember that this relationship only holds if the calorimeter does not absorb any heat from the reaction, and there is no heat exchange between the calorimeter and the outside environment.)

Next, we know that the heat absorbed by the solution depends on its specific heat, mass, and temperature change:

Equation:

$$q_{\text{solution}} = (c \times m \times \Delta T)_{\text{solution}}$$

To proceed with this calculation, we need to make a few more reasonable assumptions or approximations.

Since the solution is aqueous, we can proceed as if it were water in terms of its specific heat and mass values.

The density of water is approximately 1.0 g/mL, so 100.0 mL has a mass of about 1.0×10^2 g (two significant figures). The specific heat of water is approximately 4.184 J/g °C, so we use that for the specific heat of the solution. Substituting these values gives:

Equation:

$$q_{\text{solution}} = (4.184 \text{ J/g } ^\circ\text{C})(1.0 \times 10^2 \text{ g})(28.9 ^\circ\text{C} - 22.0 ^\circ\text{C}) = 2.9 \times 10^3 \text{ J}$$

Finally, since we are trying to find the heat of the reaction, we have:

Equation:

$$q_{\text{reaction}} = -q_{\text{solution}} = -2.9 \times 10^3 \text{ J}$$

The negative sign indicates that the reaction is exothermic. It produces 2.9 kJ of heat.

Check Your Learning

When 100 mL of 0.200 M NaCl(aq) and 100 mL of 0.200 M AgNO₃(aq), both at 21.9 °C, are mixed in a coffee cup calorimeter, the temperature increases to 23.5 °C as solid AgCl forms. How much heat is produced by this precipitation reaction? What assumptions did you make to determine your value?

Note:

Answer:

$1.34 \times 10^3 \text{ J}$; assume no heat is absorbed by the calorimeter, no heat is exchanged between the calorimeter and its surroundings, and that the specific heat and mass of the solution are the same as those for water

Note:

Thermochemistry of Hand Warmers

When working or playing outdoors on a cold day, you might use a hand warmer to warm your hands ([link](#)). A common reusable hand warmer contains a supersaturated solution of $\text{NaC}_2\text{H}_3\text{O}_2$ (sodium acetate) and a metal disc. Bending the disk creates nucleation sites around which the metastable $\text{NaC}_2\text{H}_3\text{O}_2$ quickly crystallizes (a later chapter on solutions will investigate saturation and supersaturation in more detail).

The process $\text{NaC}_2\text{H}_3\text{O}_2(aq) \rightarrow \text{NaC}_2\text{H}_3\text{O}_2(s)$ is exothermic, and the heat produced by this process is absorbed by your hands, thereby warming them (at least for a while). If the hand warmer is reheated, the $\text{NaC}_2\text{H}_3\text{O}_2$ redissolves and can be reused.



Chemical hand warmers produce heat that warms your hand on a cold day. In this one, you can see the metal disc that initiates the exothermic precipitation reaction. (credit: modification of work by Science Buddies TV/YouTube)

Another common hand warmer produces heat when it is ripped open, exposing iron and water in the hand warmer to oxygen in the air. One simplified version of this exothermic reaction is $2\text{Fe}(s) + \frac{3}{2}\text{O}_2(g) \rightarrow \text{Fe}_2\text{O}_3(s)$. Salt in the hand warmer catalyzes the reaction, so it produces heat more rapidly; cellulose, vermiculite, and activated carbon help distribute the heat evenly. Other types of hand warmers use lighter fluid (a platinum catalyst helps lighter fluid oxidize exothermically), charcoal (charcoal oxidizes in a special case), or electrical units that produce heat by passing an electrical current from a battery through resistive wires.

Note:

This [link](#) shows the precipitation reaction that occurs when the disk in a chemical hand warmer is flexed.

Example:

Heat Flow in an Instant Ice Pack

When solid ammonium nitrate dissolves in water, the solution becomes cold. This is the basis for an “instant ice pack” ([link](#)). When 3.21 g of solid NH_4NO_3 dissolves in 50.0 g of water at 24.9 °C in a calorimeter, the temperature decreases to 20.3 °C.

Calculate the value of q for this reaction and explain the meaning of its arithmetic sign. State any assumptions that you made.



An instant cold pack consists of a bag containing solid ammonium nitrate and a second bag of water. When the bag of water is broken, the pack becomes cold because the dissolution of ammonium nitrate is an endothermic process that removes thermal energy from the water. The cold pack then removes thermal energy from your body.

Solution

We assume that the calorimeter prevents heat transfer between the solution and its external environment (including the calorimeter itself), in which case:

Equation:

$$q_{\text{rxn}} = -q_{\text{soln}}$$

with “rxn” and “soln” used as shorthand for “reaction” and “solution,” respectively.

Assuming also that the specific heat of the solution is the same as that for water, we have:

Equation:

$$\begin{aligned} q_{\text{rxn}} &= -q_{\text{soln}} = -(c \times m \times \Delta T)_{\text{soln}} \\ &= -[(4.184 \text{ J/g } ^\circ\text{C}) \times (53.2 \text{ g}) \times (20.3 ^\circ\text{C} - 24.9 ^\circ\text{C})] \\ &= -[(4.184 \text{ J/g } ^\circ\text{C}) \times (53.2 \text{ g}) \times (-4.6 ^\circ\text{C})] \\ &+1.0 \times 10^3 \text{ J} = +1.0 \text{ kJ} \end{aligned}$$

The positive sign for q indicates that the dissolution is an endothermic process.

Check Your Learning

When a 3.00-g sample of KCl was added to 3.00×10^2 g of water in a coffee cup calorimeter, the temperature decreased by 1.05 °C. How much heat is involved in the dissolution of the KCl? What assumptions did you make?

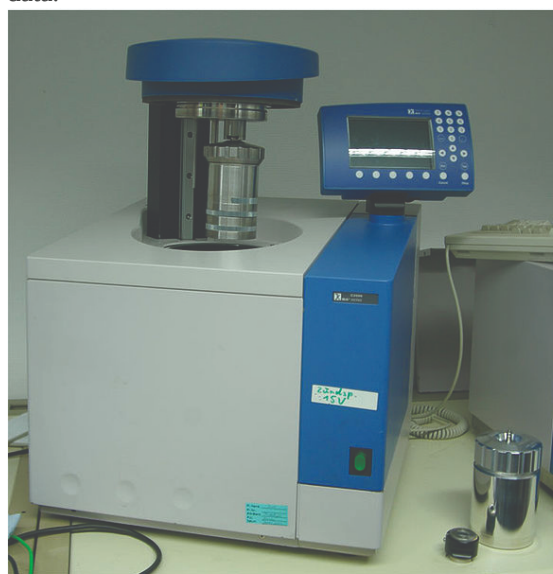
Note:

Answer:

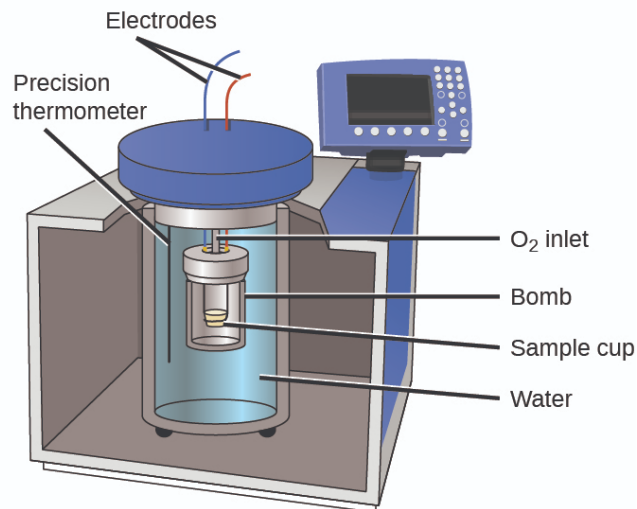
1.33 kJ; assume that the calorimeter prevents heat transfer between the solution and its external environment (including the calorimeter itself) and that the specific heat of the solution is the same as that for water

If the amount of heat absorbed by a calorimeter is too large to neglect or if we require more accurate results, then we must take into account the heat absorbed both by the solution and by the calorimeter.

The calorimeters described are designed to operate at constant (atmospheric) pressure and are convenient to measure heat flow accompanying processes that occur in solution. A different type of calorimeter that operates at constant volume, colloquially known as a **bomb calorimeter**, is used to measure the energy produced by reactions that yield large amounts of heat and gaseous products, such as combustion reactions. (The term “bomb” comes from the observation that these reactions can be vigorous enough to resemble explosions that would damage other calorimeters.) This type of calorimeter consists of a robust steel container (the “bomb”) that contains the reactants and is itself submerged in water ([link](#)). The sample is placed in the bomb, which is then filled with oxygen at high pressure. A small electrical spark is used to ignite the sample. The energy produced by the reaction is absorbed by the steel bomb and the surrounding water. The temperature increase is measured and, along with the known heat capacity of the calorimeter, is used to calculate the energy produced by the reaction. Bomb calorimeters require calibration to determine the heat capacity of the calorimeter and ensure accurate results. The calibration is accomplished using a reaction with a known q , such as a measured quantity of benzoic acid ignited by a spark from a nickel fuse wire that is weighed before and after the reaction. The temperature change produced by the known reaction is used to determine the heat capacity of the calorimeter. The calibration is generally performed each time before the calorimeter is used to gather research data.



(a)



(b)

(a) A bomb calorimeter is used to measure heat produced by reactions involving gaseous reactants or products, such as combustion. (b) The reactants are contained in the gas-tight “bomb,” which is submerged in water and surrounded by insulating materials. (credit a: modification of work by “Harbor1”/Wikimedia commons)

Note:

Click on this [link](#) to view how a bomb calorimeter is prepared for action.
This [site](#) shows calorimetric calculations using sample data.

Example:**Bomb Calorimetry**

When 3.12 g of glucose, $\text{C}_6\text{H}_{12}\text{O}_6$, is burned in a bomb calorimeter, the temperature of the calorimeter increases from 23.8 °C to 35.6 °C. The calorimeter contains 775 g of water, and the bomb itself has a heat capacity of 893 J/°C. How much heat was produced by the combustion of the glucose sample?

Solution

The combustion produces heat that is primarily absorbed by the water and the bomb. (The amounts of heat absorbed by the reaction products and the unreacted excess oxygen are relatively small and dealing with them is beyond the scope of this text. We will neglect them in our calculations.)

The heat produced by the reaction is absorbed by the water and the bomb:

Equation:

$$\begin{aligned}q_{\text{rxn}} &= -(q_{\text{water}} + q_{\text{bomb}}) \\&= -[(4.184 \text{ J/g } ^\circ\text{C}) \times (775 \text{ g}) \times (35.6 ^\circ\text{C} - 23.8 ^\circ\text{C}) + 893 \text{ J/}^\circ\text{C} \times (35.6 ^\circ\text{C} - 23.8 ^\circ\text{C})] \\&= -(38,300 \text{ J} + 10,500 \text{ J}) \\&= -48,800 \text{ J} = -48.8 \text{ kJ}\end{aligned}$$

This reaction released 48.7 kJ of heat when 3.12 g of glucose was burned.

Check Your Learning

When 0.963 g of benzene, C_6H_6 , is burned in a bomb calorimeter, the temperature of the calorimeter increases by 8.39 °C. The bomb has a heat capacity of 784 J/°C and is submerged in 925 mL of water. How much heat was produced by the combustion of the glucose sample?

Note:**Answer:**

39.0 kJ

Since the first one was constructed in 1899, 35 calorimeters have been built to measure the heat produced by a living person.^[footnote] These whole-body calorimeters of various designs are large enough to hold an individual human being. More recently, whole-room calorimeters allow for relatively normal activities to be performed, and these calorimeters generate data that more closely reflect the real world. These calorimeters are used to measure the metabolism of individuals under different environmental conditions, different dietary regimes, and with different health conditions, such as diabetes. In humans, metabolism is typically measured in Calories per day. A **nutritional calorie (Calorie)** is the energy unit used to quantify the amount of energy derived from the metabolism of foods; one Calorie is equal to 1000 calories (1 kcal), the amount of energy needed to heat 1 kg of water by 1 °C.

Francis D. Reardon et al. "The Snellen human calorimeter revisited, re-engineered and upgraded: Design and performance characteristics." *Medical and Biological Engineering and Computing* 8 (2006)721–28, <http://link.springer.com/article/10.1007/s11517-006-0086-5>.

Note:**Measuring Nutritional Calories**

In your day-to-day life, you may be more familiar with energy being given in Calories, or nutritional calories, which are used to quantify the amount of energy in foods. One calorie (cal) = exactly 4.184 joules, and one Calorie (note the capitalization) = 1000 cal, or 1 kcal. (This is approximately the amount of energy needed to heat 1 kg of water by 1 °C.)

The macronutrients in food are proteins, carbohydrates, and fats or oils. Proteins provide about 4 Calories per gram, carbohydrates also provide about 4 Calories per gram, and fats and oils provide about 9 Calories/g.

Nutritional labels on food packages show the caloric content of one serving of the food, as well as the breakdown into Calories from each of the three macronutrients ([link](#)).



(a)

Sample label for macaroni & cheese

Nutrition Facts	
Serving Size 1 cup (228g) Servings Per Container 2	
Amount Per Serving	
Calories 250	Calories from Fat 110
% Daily Value*	
Total Fat 12g	18%
Saturated Fat 3g	15%
Trans Fat 3g	
Cholesterol 30mg	10%
Sodium 470mg	20%
Total Carbohydrates 31g	10%
Dietary Fiber 0g	0%
Sugars 5g	
Protein 5g	
Vitamin A	4%
Vitamin C	2%
Calcium	20%
Iron	4%

*Percent Daily Values are based on a 2,000 calorie diet.
Your Daily Values may be higher or lower depending on your calorie needs.

	Calories	2,000	2,500
Total Fat	Less than	65g	80g
Sat Fat	Less than	20g	25g
Cholesterol	Less than	300mg	300mg
Sodium	Less than	2,400mg	2,400mg
Total Carbohydrate		300g	375g
Dietary Fiber		25g	30g

① Start here →
② Check Calories
③ Limit these nutrients
④ Get enough of these nutrients
⑤ Footnote
⑥ Quick guide to % DV
+ 5% or less is low
+ 20% or more is high

(b)

(a) Macaroni and cheese contain energy in the form of the macronutrients in the food. (b) The food's nutritional information is shown on the package label. In the US, the energy content is given in Calories (per serving); the rest of the world usually uses kilojoules. (credit a: modification of work by "Rex Roof"/Flickr)

For the example shown in (b), the total energy per 228-g portion is calculated by:

Equation:

$$(5 \text{ g protein} \times 4 \text{ Calories/g}) + (31 \text{ g carb} \times 4 \text{ Calories/g}) + (12 \text{ g fat} \times 9 \text{ Calories/g}) = 252 \text{ Calories}$$

So, you can use food labels to count your Calories. But where do the values come from? And how accurate are they? The caloric content of foods can be determined by using bomb calorimetry; that is, by burning the food and measuring the energy it contains. A sample of food is weighed, mixed in a blender, freeze-dried, ground into powder, and formed into a pellet. The pellet is burned inside a bomb calorimeter, and the measured temperature change is converted into energy per gram of food.

Today, the caloric content on food labels is derived using a method called the Atwater system that uses the average caloric content of the different chemical constituents of food, protein, carbohydrate, and fats. The average amounts are those given in the equation and are derived from the various results given by bomb calorimetry of whole foods. The carbohydrate amount is discounted a certain amount for the fiber content, which is indigestible carbohydrate. To determine the energy content of a food, the quantities of carbohydrate,

protein, and fat are each multiplied by the average Calories per gram for each and the products summed to obtain the total energy.

Note:

Click on this [link](#) to access the US Department of Agriculture (USDA) National Nutrient Database, containing nutritional information on over 8000 foods.

Key Concepts and Summary

Calorimetry is used to measure the amount of thermal energy transferred in a chemical or physical process. This requires careful measurement of the temperature change that occurs during the process and the masses of the system and surroundings. These measured quantities are then used to compute the amount of heat produced or consumed in the process using known mathematical relations.

Calorimeters are designed to minimize energy exchange between their contents and the external environment. They range from simple coffee cup calorimeters used by introductory chemistry students to sophisticated bomb calorimeters used to determine the energy content of food.

Chemistry End of Chapter Exercises

Exercise:

Problem:

A 500-mL bottle of water at room temperature and a 2-L bottle of water at the same temperature were placed in a refrigerator. After 30 minutes, the 500-mL bottle of water had cooled to the temperature of the refrigerator. An hour later, the 2-L of water had cooled to the same temperature. When asked which sample of water lost the most heat, one student replied that both bottles lost the same amount of heat because they started at the same temperature and finished at the same temperature. A second student thought that the 2-L bottle of water lost more heat because there was more water. A third student believed that the 500-mL bottle of water lost more heat because it cooled more quickly. A fourth student thought that it was not possible to tell because we do not know the initial temperature and the final temperature of the water. Indicate which of these answers is correct and describe the error in each of the other answers.

Exercise:

Problem:

Would the amount of heat measured for the reaction in [\[link\]](#) be greater, lesser, or remain the same if we used a calorimeter that was a poorer insulator than a coffee cup calorimeter? Explain your answer.

Solution:

lesser; more heat would be lost to the coffee cup and the environment and so ΔT for the water would be lesser and the calculated q would be lesser

Exercise:

Problem:

Would the amount of heat absorbed by the dissolution in [\[link\]](#) appear greater, lesser, or remain the same if the experimenter used a calorimeter that was a poorer insulator than a coffee cup calorimeter? Explain your answer.

Exercise:**Problem:**

Would the amount of heat absorbed by the dissolution in [\[link\]](#) appear greater, lesser, or remain the same if the heat capacity of the calorimeter were taken into account? Explain your answer.

Solution:

greater, since taking the calorimeter's heat capacity into account will compensate for the thermal energy transferred to the solution from the calorimeter; this approach includes the calorimeter itself, along with the solution, as "surroundings": $q_{\text{rxn}} = -(q_{\text{solution}} + q_{\text{calorimeter}})$; since both q_{solution} and $q_{\text{calorimeter}}$ are negative, including the latter term (q_{rxn}) will yield a greater value for the heat of the dissolution

Exercise:**Problem:**

How many milliliters of water at 23 °C with a density of 1.00 g/mL must be mixed with 180 mL (about 6 oz) of coffee at 95 °C so that the resulting combination will have a temperature of 60 °C? Assume that coffee and water have the same density and the same specific heat.

Exercise:**Problem:**

How much will the temperature of a cup (180 g) of coffee at 95 °C be reduced when a 45 g silver spoon (specific heat 0.24 J/g °C) at 25 °C is placed in the coffee and the two are allowed to reach the same temperature? Assume that the coffee has the same density and specific heat as water.

Solution:

The temperature of the coffee will drop 1 degree.

Exercise:**Problem:**

A 45-g aluminum spoon (specific heat 0.88 J/g °C) at 24 °C is placed in 180 mL (180 g) of coffee at 85 °C and the temperature of the two become equal.

(a) What is the final temperature when the two become equal? Assume that coffee has the same specific heat as water.

(b) The first time a student solved this problem she got an answer of 88 °C. Explain why this is clearly an incorrect answer.

Exercise:**Problem:**

The temperature of the cooling water as it leaves the hot engine of an automobile is 240 °F. After it passes through the radiator it has a temperature of 175 °F. Calculate the amount of heat transferred from the engine to the surroundings by one gallon of water with a specific heat of 4.184 J/g °C.

Solution:

$$5.7 \times 10^2 \text{ kJ}$$

Exercise:

Problem:

A 70.0-g piece of metal at 80.0 °C is placed in 100 g of water at 22.0 °C contained in a calorimeter like that shown in [\[link\]](#). The metal and water come to the same temperature at 24.6 °C. How much heat did the metal give up to the water? What is the specific heat of the metal?

Exercise:**Problem:**

If a reaction produces 1.506 kJ of heat, which is trapped in 30.0 g of water initially at 26.5 °C in a calorimeter like that in [\[link\]](#), what is the resulting temperature of the water?

Solution:

38.5 °C

Exercise:**Problem:**

A 0.500-g sample of KCl is added to 50.0 g of water in a calorimeter ([\[link\]](#)). If the temperature decreases by 1.05 °C, what is the approximate amount of heat involved in the dissolution of the KCl, assuming the specific heat of the resulting solution is 4.18 J/g °C? Is the reaction exothermic or endothermic?

Exercise:**Problem:**

Dissolving 3.0 g of CaCl₂(s) in 150.0 g of water in a calorimeter ([\[link\]](#)) at 22.4 °C causes the temperature to rise to 25.8 °C. What is the approximate amount of heat involved in the dissolution, assuming the specific heat of the resulting solution is 4.18 J/g °C? Is the reaction exothermic or endothermic?

Solution:

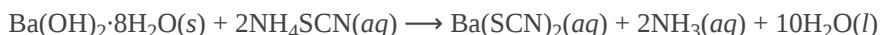
−2.2 kJ; The heat produced shows that the reaction is exothermic.

Exercise:**Problem:**

When 50.0 g of 0.200 M NaCl(aq) at 24.1 °C is added to 100.0 g of 0.100 M AgNO₃(aq) at 24.1 °C in a calorimeter, the temperature increases to 25.2 °C as AgCl(s) forms. Assuming the specific heat of the solution and products is 4.20 J/g °C, calculate the approximate amount of heat in joules produced.

Exercise:**Problem:**

The addition of 3.15 g of Ba(OH)₂·8H₂O to a solution of 1.52 g of NH₄SCN in 100 g of water in a calorimeter caused the temperature to fall by 3.1 °C. Assuming the specific heat of the solution and products is 4.20 J/g °C, calculate the approximate amount of heat absorbed by the reaction, which can be represented by the following equation:



Solution:

1.4 kJ

Exercise:

Problem:

The reaction of 50 mL of acid and 50 mL of base described in [\[link\]](#) increased the temperature of the solution by 6.9 degrees C. How much would the temperature have increased if 100 mL of acid and 100 mL of base had been used in the same calorimeter starting at the same temperature of 22.0 °C? Explain your answer.

Exercise:**Problem:**

If the 3.21 g of NH_4NO_3 in [\[link\]](#) were dissolved in 100.0 g of water under the same conditions, how much would the temperature change? Explain your answer.

Solution:

22.6. Since the mass and the heat capacity of the solution is approximately equal to that of the water, the two-fold increase in the amount of water leads to a two-fold decrease of the temperature change.

Exercise:**Problem:**

When 1.0 g of fructose, $\text{C}_6\text{H}_{12}\text{O}_6(\text{s})$, a sugar commonly found in fruits, is burned in oxygen in a bomb calorimeter, the temperature of the calorimeter increases by 1.58 °C. If the heat capacity of the calorimeter and its contents is 9.90 kJ/°C, what is q for this combustion?

Exercise:**Problem:**

When a 0.740-g sample of trinitrotoluene (TNT), $\text{C}_7\text{H}_5\text{N}_2\text{O}_6$, is burned in a bomb calorimeter, the temperature increases from 23.4 °C to 26.9 °C. The heat capacity of the calorimeter is 534 J/°C, and it contains 675 mL of water. How much heat was produced by the combustion of the TNT sample?

Solution:

11.7 kJ

Exercise:**Problem:**

One method of generating electricity is by burning coal to heat water, which produces steam that drives an electric generator. To determine the rate at which coal is to be fed into the burner in this type of plant, the heat of combustion per ton of coal must be determined using a bomb calorimeter. When 1.00 g of coal is burned in a bomb calorimeter ([\[link\]](#)), the temperature increases by 1.48 °C. If the heat capacity of the calorimeter is 21.6 kJ/°C, determine the heat produced by combustion of a ton of coal (2.000×10^3 pounds).

Exercise:**Problem:**

The amount of fat recommended for someone with a daily diet of 2000 Calories is 65 g. What percent of the calories in this diet would be supplied by this amount of fat if the average number of Calories for fat is 9.1 Calories/g?

Solution:

30%

Exercise:**Problem:**

A teaspoon of the carbohydrate sucrose (common sugar) contains 16 Calories (16 kcal). What is the mass of one teaspoon of sucrose if the average number of Calories for carbohydrates is 4.1 Calories/g?

Exercise:**Problem:**

What is the maximum mass of carbohydrate in a 6-oz serving of diet soda that contains less than 1 Calorie per can if the average number of Calories for carbohydrates is 4.1 Calories/g?

Solution:

0.24 g

Exercise:**Problem:**

A pint of premium ice cream can contain 1100 Calories. What mass of fat, in grams and pounds, must be produced in the body to store an extra 1.1×10^3 Calories if the average number of Calories for fat is 9.1 Calories/g?

Exercise:**Problem:**

A serving of a breakfast cereal contains 3 g of protein, 18 g of carbohydrates, and 6 g of fat. What is the Calorie content of a serving of this cereal if the average number of Calories for fat is 9.1 Calories/g, for carbohydrates is 4.1 Calories/g, and for protein is 4.1 Calories/g?

Solution:

1.4×10^2 Calories

Exercise:**Problem:**

Which is the least expensive source of energy in kilojoules per dollar: a box of breakfast cereal that weighs 32 ounces and costs \$4.23, or a liter of isooctane (density, 0.6919 g/mL) that costs \$0.45? Compare the nutritional value of the cereal with the heat produced by combustion of the isooctane under standard conditions. A 1.0-ounce serving of the cereal provides 130 Calories.

Glossary

bomb calorimeter

device designed to measure the energy change for processes occurring under conditions of constant volume; commonly used for reactions involving solid and gaseous reactants or products

calorimeter

device used to measure the amount of heat absorbed or released in a chemical or physical process

calorimetry

process of measuring the amount of heat involved in a chemical or physical process

nutritional calorie (Calorie)

unit used for quantifying energy provided by digestion of foods, defined as 1000 cal or 1 kcal

surroundings

all matter other than the system being studied

system

portion of matter undergoing a chemical or physical change being studied

Enthalpy

By the end of this section, you will be able to:

- State the first law of thermodynamics
- Define enthalpy and explain its classification as a state function
- Write and balance thermochemical equations
- Calculate enthalpy changes for various chemical reactions
- Explain Hess's law and use it to compute reaction enthalpies

Thermochemistry is a branch of **chemical thermodynamics**, the science that deals with the relationships between heat, work, and other forms of energy in the context of chemical and physical processes. As we concentrate on thermochemistry in this chapter, we need to consider some widely used concepts of thermodynamics.

Substances act as reservoirs of energy, meaning that energy can be added to them or removed from them. Energy is stored in a substance when the kinetic energy of its atoms or molecules is raised. The greater kinetic energy may be in the form of increased translations (travel or straight-line motions), vibrations, or rotations of the atoms or molecules. When thermal energy is lost, the intensities of these motions decrease and the kinetic energy falls. The total of all possible kinds of energy present in a substance is called the **internal energy (U)**, sometimes symbolized as E .

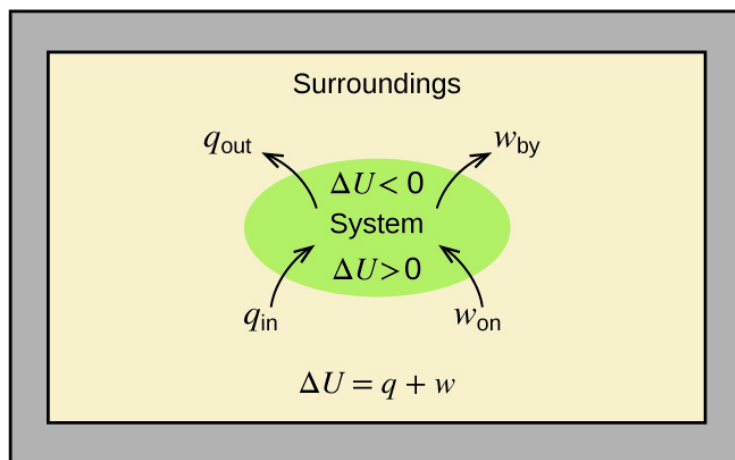
As a system undergoes a change, its internal energy can change, and energy can be transferred from the system to the surroundings, or from the surroundings to the system. Energy is transferred into a system when it absorbs heat (q) from the surroundings or when the surroundings do work (w) on the system. For example, energy is transferred into room-temperature metal wire if it is immersed in hot water (the wire absorbs heat from the water), or if you rapidly bend the wire back and forth (the wire becomes warmer because of the work done on it). Both processes increase the internal energy of the wire, which is reflected in an increase in the wire's temperature. Conversely, energy is transferred out of a system when heat is lost from the system, or when the system does work on the surroundings.

The relationship between internal energy, heat, and work can be represented by the equation:

Equation:

$$\Delta U = q + w$$

as shown in [\[link\]](#). This is one version of the **first law of thermodynamics**, and it shows that the internal energy of a system changes through heat flow into or out of the system (positive q is heat flow in; negative q is heat flow out) or work done on or by the system. The work, w , is positive if it is done on the system and negative if it is done by the system.



The internal energy, U , of a system can be changed by heat

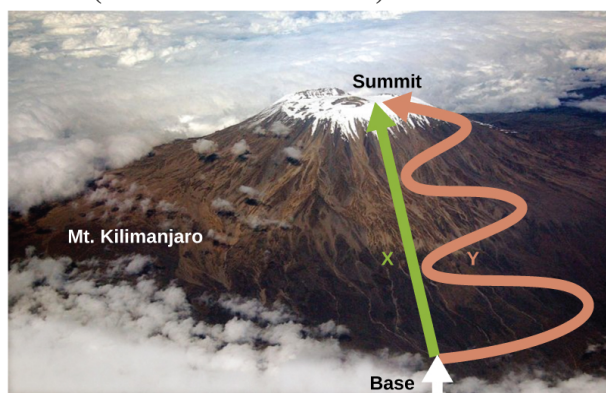
flow and work. If heat flows into the system, q_{in} , or work is done on the system, w_{on} , its internal energy increases, $\Delta U > 0$. If heat flows out of the system, q_{out} , or work is done by the system, w_{by} , its internal energy decreases, $\Delta U < 0$.

A type of work called **expansion work** (or pressure-volume work) occurs when a system pushes back the surroundings against a restraining pressure, or when the surroundings compress the system. An example of this occurs during the operation of an internal combustion engine. The reaction of gasoline and oxygen is exothermic. Some of this energy is given off as heat, and some does work pushing the piston in the cylinder. The substances involved in the reaction are the system, and the engine and the rest of the universe are the surroundings. The system loses energy by both heating and doing work on the surroundings, and its internal energy decreases. (The engine is able to keep the car moving because this process is repeated many times per second while the engine is running.) We will consider how to determine the amount of work involved in a chemical or physical change in the chapter on thermodynamics.

Note:

This view of [an internal combustion engine](#) illustrates the conversion of energy produced by the exothermic combustion reaction of a fuel such as gasoline into energy of motion.

As discussed, the relationship between internal energy, heat, and work can be represented as $\Delta U = q + w$. Internal energy is an example of a **state function** (or state variable), whereas heat and work are not state functions. The value of a state function depends only on the state that a system is in, and not on how that state is reached. If a quantity is not a state function, then its value *does* depend on how the state is reached. An example of a state function is altitude or elevation. If you stand on the summit of Mt. Kilimanjaro, you are at an altitude of 5895 m, and it does not matter whether you hiked there or parachuted there. The distance you traveled to the top of Kilimanjaro, however, is not a state function. You could climb to the summit by a direct route or by a more roundabout, circuitous path ([link](#)). The distances traveled would differ (distance is not a state function) but the elevation reached would be the same (altitude is a state function).



Paths X and Y represent two different routes to the summit of Mt. Kilimanjaro. Both have the same change in elevation (altitude or elevation on a mountain is a state function; it does not depend on path), but they have very different distances traveled (distance walked is not a state function; it depends on the path). (credit: modification of work by Paul Shaffner)

Chemists ordinarily use a property known as **enthalpy (H)** to describe the thermodynamics of chemical and physical processes. Enthalpy is defined as the sum of a system's internal energy (U) and the mathematical product of its pressure (P) and volume (V):

Equation:

$$H = U + PV$$

Enthalpy is also a state function. Enthalpy values for specific substances cannot be measured directly; only enthalpy *changes* for chemical or physical processes can be determined. For processes that take place at constant pressure (a common condition for many chemical and physical changes), the **enthalpy change (ΔH)** is:

Equation:

$$\Delta H = \Delta U + P\Delta V$$

The mathematical product $P\Delta V$ represents work (w), namely, expansion or pressure-volume work as noted. By their definitions, the arithmetic signs of ΔV and w will always be opposite:

Equation:

$$P\Delta V = -w$$

Substituting this equation and the definition of internal energy into the enthalpy-change equation yields:

Equation:

$$\begin{aligned}\Delta H &= \Delta U + P\Delta V \\ &= q_p + w - w \\ &= q_p\end{aligned}$$

where q_p is the heat of reaction under conditions of constant pressure.

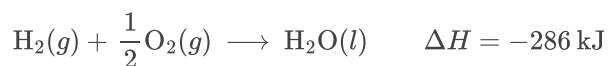
And so, if a chemical or physical process is carried out at constant pressure with the only work done caused by expansion or contraction, then the heat flow (q_p) and enthalpy change (ΔH) for the process are equal.

The heat given off when you operate a Bunsen burner is equal to the enthalpy change of the methane combustion reaction that takes place, since it occurs at the essentially constant pressure of the atmosphere. On the other hand, the heat produced by a reaction measured in a bomb calorimeter ([link](#)) is not equal to ΔH because the closed, constant-volume metal container prevents the pressure from remaining constant (it may increase or decrease if the reaction yields increased or decreased amounts of gaseous species). Chemists usually perform experiments under normal atmospheric conditions, at constant external pressure with $q = \Delta H$, which makes enthalpy the most convenient choice for determining heat changes for chemical reactions.

The following conventions apply when using ΔH :

- A negative value of an enthalpy change, $\Delta H < 0$, indicates an exothermic reaction; a positive value, $\Delta H > 0$, indicates an endothermic reaction. If the direction of a chemical equation is reversed, the arithmetic sign of its ΔH is changed (a process that is endothermic in one direction is exothermic in the opposite direction).
- Chemists use a thermochemical equation to represent the changes in both matter and energy. In a thermochemical equation, the enthalpy change of a reaction is shown as a ΔH value following the equation for the reaction. This ΔH value indicates the amount of heat associated with the reaction involving the number of moles of reactants and products *as shown in the chemical equation*. For example, consider this equation:

Equation:



This equation indicates that when 1 mole of hydrogen gas and $\frac{1}{2}$ mole of oxygen gas at some temperature and pressure change to 1 mole of liquid water at the same temperature and pressure, 286 kJ of heat are released to the surroundings. If the coefficients of the chemical equation are multiplied by some factor, the enthalpy change must be multiplied by that same factor (ΔH is an extensive property):

Equation:

(two-fold increase in amounts)

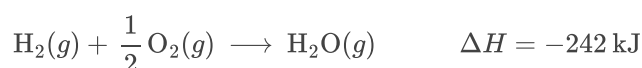


(two-fold decrease in amounts)



- The enthalpy change of a reaction depends on the physical states of the reactants and products, so these must be shown. For example, when 1 mole of hydrogen gas and $\frac{1}{2}$ mole of oxygen gas change to 1 mole of liquid water at the same temperature and pressure, 286 kJ of heat are released. If gaseous water forms, only 242 kJ of heat are released.

Equation:

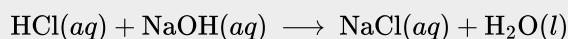


Example:

Writing Thermochemical Equations

When 0.0500 mol of $\text{HCl}(aq)$ reacts with 0.0500 mol of $\text{NaOH}(aq)$ to form 0.0500 mol of $\text{NaCl}(aq)$, 2.9 kJ of heat are produced. Write a balanced thermochemical equation for the reaction of one mole of HCl .

Equation:



Solution

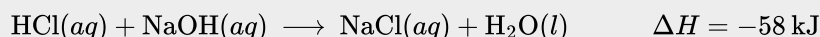
For the reaction of 0.0500 mol acid (HCl), $q = -2.9 \text{ kJ}$. The reactants are provided in stoichiometric amounts (same molar ratio as in the balanced equation), and so the amount of acid may be used to calculate a molar enthalpy change. Since ΔH is an extensive property, it is proportional to the amount of acid neutralized:

Equation:

$$\Delta H = 1 \cancel{\text{mol HCl}} \times \frac{-2.9 \text{ kJ}}{0.0500 \cancel{\text{mol HCl}}} = -58 \text{ kJ}$$

The thermochemical equation is then

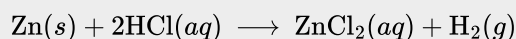
Equation:



Check Your Learning

When 1.34 g $\text{Zn}(s)$ reacts with 60.0 mL of 0.750 M $\text{HCl}(aq)$, 3.14 kJ of heat are produced. Determine the enthalpy change per mole of zinc reacting for the reaction:

Equation:



Note:

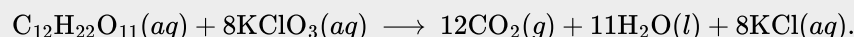
Answer:

$$\Delta H = -153 \text{ kJ}$$

Be sure to take both stoichiometry and limiting reactants into account when determining the ΔH for a chemical reaction.

Example:**Writing Thermochemical Equations**

A gummy bear contains 2.67 g sucrose, $\text{C}_{12}\text{H}_{22}\text{O}_{11}$. When it reacts with 7.19 g potassium chlorate, KClO_3 , 43.7 kJ of heat are produced. Write a thermochemical equation for the reaction of one mole of sucrose:

Equation:**Solution**

Unlike the previous example exercise, this one does not involve the reaction of stoichiometric amounts of reactants, and so the *limiting reactant* must be identified (it limits the yield of the reaction and the amount of thermal energy produced or consumed).

The provided amounts of the two reactants are

Equation:

$$\begin{aligned}(2.67 \text{ g}) (1 \text{ mol}/342.3 \text{ g}) &= 0.00780 \text{ mol C}_{12}\text{H}_{22}\text{O}_{11} \\ (7.19 \text{ g}) (1 \text{ mol}/122.5 \text{ g}) &= 0.0587 \text{ mol KClO}_3\end{aligned}$$

The provided molar ratio of perchlorate-to-sucrose is then

Equation:

$$0.0587 \text{ mol KClO}_3 / 0.00780 \text{ mol C}_{12}\text{H}_{22}\text{O}_{11} = 7.52$$

The balanced equation indicates 8 mol KClO_3 are required for reaction with 1 mol $\text{C}_{12}\text{H}_{22}\text{O}_{11}$. Since the provided amount of KClO_3 is less than the stoichiometric amount, it is the limiting reactant and may be used to compute the enthalpy change:

Equation:

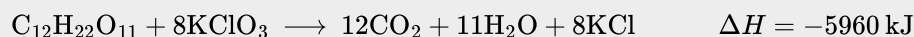
$$\Delta H = -43.7 \text{ kJ} / 0.0587 \text{ mol KClO}_3 = 744 \text{ kJ/mol KClO}_3$$

Because the equation, as written, represents the reaction of 8 mol KClO_3 , the enthalpy change is

Equation:

$$(744 \text{ kJ/mol KClO}_3) (8 \text{ mol KClO}_3) = 5960 \text{ kJ}$$

The enthalpy change for this reaction is -5960 kJ , and the thermochemical equation is:

Equation:**Check Your Learning**

When 1.42 g of iron reacts with 1.80 g of chlorine, 3.22 g of $\text{FeCl}_2(s)$ and 8.60 kJ of heat is produced. What is the enthalpy change for the reaction when 1 mole of $\text{FeCl}_2(s)$ is produced?

Note:

Answer:

$$\Delta H = -338 \text{ kJ}$$

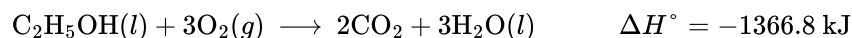
Enthalpy changes are typically tabulated for reactions in which both the reactants and products are at the same conditions. A **standard state** is a commonly accepted set of conditions used as a reference point for the determination of properties under other different conditions. For chemists, the IUPAC standard state refers to materials under a pressure of 1 bar and solutions at 1 M, and does not specify a temperature. Many thermochemical tables list values with a standard state of 1 atm. Because the ΔH of a reaction changes very little with such small changes in pressure (1 bar = 0.987 atm), ΔH values (except for the most precisely measured values) are essentially the same under both sets of standard conditions. We will include a superscripted “o” in the enthalpy change symbol to designate standard state. Since the usual (but not technically standard) temperature is 298.15 K, this temperature will be assumed unless some other temperature is specified. Thus, the symbol (ΔH°) is used to indicate an enthalpy change for a process occurring under these conditions. (The symbol ΔH is used to indicate an enthalpy change for a reaction occurring under nonstandard conditions.)

The enthalpy changes for many types of chemical and physical processes are available in the reference literature, including those for combustion reactions, phase transitions, and formation reactions. As we discuss these quantities, it is important to pay attention to the *extensive* nature of enthalpy and enthalpy changes. Since the enthalpy change for a given reaction is proportional to the amounts of substances involved, it may be reported on that basis (i.e., as the ΔH for specific amounts of reactants). However, we often find it more useful to divide one extensive property (ΔH) by another (amount of substance), and report a per-amount *intensive* value of ΔH , often “normalized” to a per-mole basis. (Note that this is similar to determining the intensive property specific heat from the extensive property heat capacity, as seen previously.)

Standard Enthalpy of Combustion

Standard enthalpy of combustion (ΔH_C°) is the enthalpy change when 1 mole of a substance burns (combines vigorously with oxygen) under standard state conditions; it is sometimes called “heat of combustion.” For example, the enthalpy of combustion of ethanol, -1366.8 kJ/mol , is the amount of heat produced when one mole of ethanol undergoes complete combustion at 25°C and 1 atmosphere pressure, yielding products also at 25°C and 1 atm.

Equation:



Enthalpies of combustion for many substances have been measured; a few of these are listed in [\[link\]](#). Many readily available substances with large enthalpies of combustion are used as fuels, including hydrogen, carbon (as coal or charcoal), and **hydrocarbons** (compounds containing only hydrogen and carbon), such as methane, propane, and the major components of gasoline.

Standard Molar Enthalpies of Combustion

Standard Molar	Enthalpies of Combustion	Enthalpy of Combustion,
Substance	Combustion Reaction	ΔH_c° ($\frac{\text{kJ}}{\text{mol}}$ at 25 °C)

Substance	Combustion Reaction	Enthalpy of Combustion, ΔH_c° ($\frac{\text{kJ}}{\text{mol}}$ at 25 °C)
carbon	$\text{C}(s) + \text{O}_2(g) \longrightarrow \text{CO}_2(g)$	-393.5
hydrogen	$\text{H}_2(g) + \frac{1}{2} \text{O}_2(g) \longrightarrow \text{H}_2\text{O}(l)$	-285.8
magnesium	$\text{Mg}(s) + \frac{1}{2} \text{O}_2(g) \longrightarrow \text{MgO}(s)$	-601.6
sulfur	$\text{S}(s) + \text{O}_2(g) \longrightarrow \text{SO}_2(g)$	-296.8
carbon monoxide	$\text{CO}(g) + \frac{1}{2} \text{O}_2(g) \longrightarrow \text{CO}_2(g)$	-283.0
methane	$\text{CH}_4(g) + 2\text{O}_2(g) \longrightarrow \text{CO}_2(g) + 2\text{H}_2\text{O}(l)$	-890.8
acetylene	$\text{C}_2\text{H}_2(g) + \frac{5}{2} \text{O}_2(g) \longrightarrow 2\text{CO}_2(g) + \text{H}_2\text{O}(l)$	-1301.1
ethanol	$\text{C}_2\text{H}_5\text{OH}(l) + 3\text{O}_2(g) \longrightarrow 2\text{CO}_2(g) + 3\text{H}_2\text{O}(l)$	-1366.8
methanol	$\text{CH}_3\text{OH}(l) + \frac{3}{2} \text{O}_2(g) \longrightarrow \text{CO}_2(g) + 2\text{H}_2\text{O}(l)$	-726.1
isooctane	$\text{C}_8\text{H}_{18}(l) + \frac{25}{2} \text{O}_2(g) \longrightarrow 8\text{CO}_2(g) + 9\text{H}_2\text{O}(l)$	-5461

Example:

Using Enthalpy of Combustion

As [\[link\]](#) suggests, the combustion of gasoline is a highly exothermic process. Let us determine the approximate amount of heat produced by burning 1.00 L of gasoline, assuming the enthalpy of combustion of gasoline is the same as that of isooctane, a common component of gasoline. The density of isooctane is 0.692 g/mL.



The combustion of gasoline is very exothermic. (credit: modification of work by “AlexEagle”/Flickr)

Solution

Starting with a known amount (1.00 L of isooctane), we can perform conversions between units until we arrive at the desired amount of heat or energy. The enthalpy of combustion of isooctane provides one of the necessary conversions. [\[link\]](#) gives this value as -5460 kJ per 1 mole of isooctane (C_8H_{18}).

Using these data,

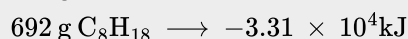
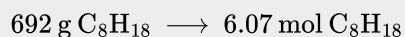
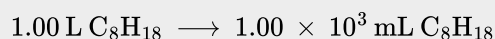
Equation:

$$1.00 \text{ L C}_8\text{H}_{18} \times \frac{1000 \text{ mL C}_8\text{H}_{18}}{1 \text{ L C}_8\text{H}_{18}} \times \frac{0.692 \text{ g C}_8\text{H}_{18}}{1 \text{ mL C}_8\text{H}_{18}} \times \frac{1 \text{ mol C}_8\text{H}_{18}}{114 \text{ g C}_8\text{H}_{18}} \times \frac{-5460 \text{ kJ}}{1 \text{ mol C}_8\text{H}_{18}} = -3.31 \times 10^4 \text{ kJ}$$

The combustion of 1.00 L of isooctane produces 33,100 kJ of heat. (This amount of energy is enough to melt 99.2 kg, or about 218 lbs, of ice.)

Note: If you do this calculation one step at a time, you would find:

Equation:



Check Your Learning

How much heat is produced by the combustion of 125 g of acetylene?

Note:

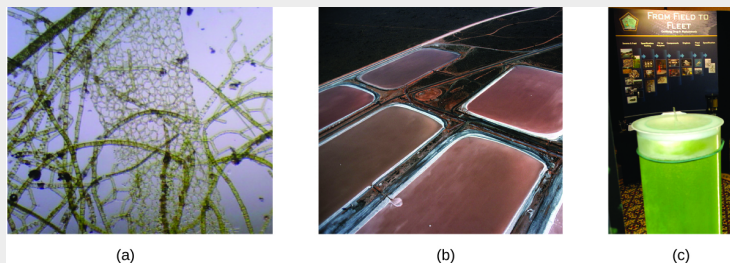
Answer:

$$6.25 \times 10^3 \text{ kJ}$$

Note:

Emerging Algae-Based Energy Technologies (Biofuels)

As reserves of fossil fuels diminish and become more costly to extract, the search is ongoing for replacement fuel sources for the future. Among the most promising biofuels are those derived from algae ([link](#)). The species of algae used are nontoxic, biodegradable, and among the world's fastest growing organisms. About 50% of algal weight is oil, which can be readily converted into fuel such as biodiesel. Algae can yield 26,000 gallons of biofuel per hectare—much more energy per acre than other crops. Some strains of algae can flourish in brackish water that is not usable for growing other crops. Algae can produce biodiesel, biogasoline, ethanol, butanol, methane, and even jet fuel.



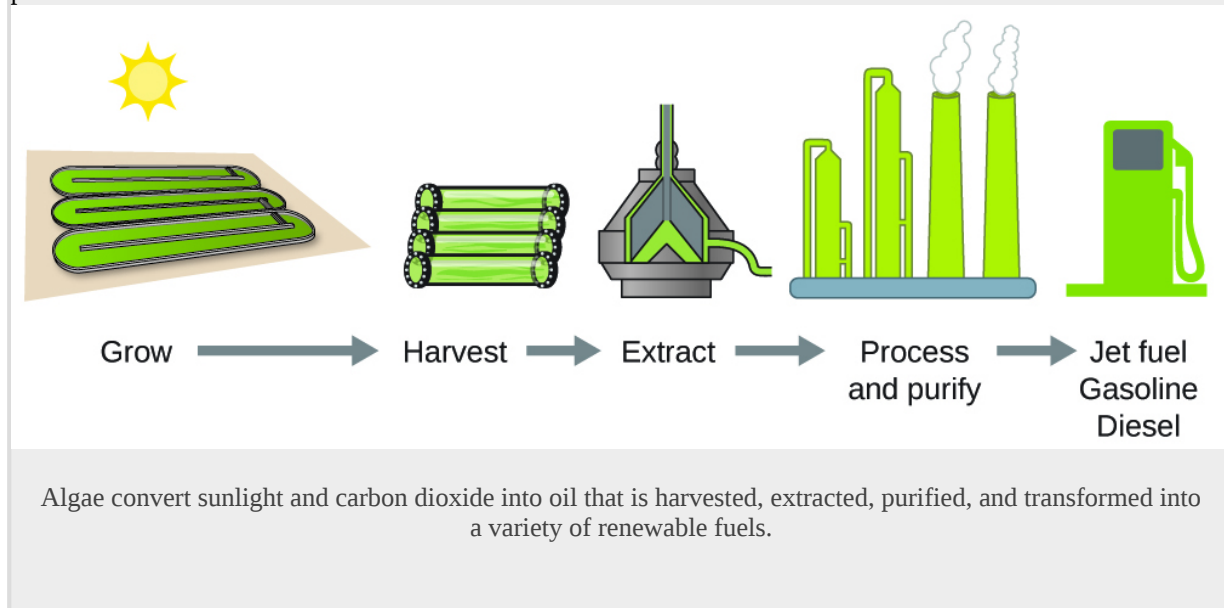
(a) Tiny algal organisms can be (b) grown in large quantities and eventually (c) turned into a useful fuel such as biodiesel.

(credit a: modification of work by Micah Sittig; credit b: modification of work by Robert Kerton; credit c: modification of work by John F. Williams)

According to the US Department of Energy, only 39,000 square kilometers (about 0.4% of the land mass of the US or less than $\frac{1}{7}$ of the area used to grow corn) can produce enough algal fuel to replace all the petroleum-based

fuel used in the US. The cost of algal fuels is becoming more competitive—for instance, the US Air Force is producing jet fuel from algae at a total cost of under \$5 per gallon.^[footnote] The process used to produce algal fuel is as follows: grow the algae (which use sunlight as their energy source and CO₂ as a raw material); harvest the algae; extract the fuel compounds (or precursor compounds); process as necessary (e.g., perform a transesterification reaction to make biodiesel); purify; and distribute (^[link]).

For more on algal fuel, see <http://www.theguardian.com/environment/2010/feb/13/algae-solve-pentagon-fuel-problem>.



Note:

Click [here](#) to learn more about the process of creating algae biofuel.

Standard Enthalpy of Formation

A **standard enthalpy of formation** ΔH_f° is an enthalpy change for a reaction in which exactly 1 mole of a pure substance is formed from free elements in their most stable states under standard state conditions. These values are especially useful for computing or predicting enthalpy changes for chemical reactions that are impractical or dangerous to carry out, or for processes for which it is difficult to make measurements. If we have values for the appropriate standard enthalpies of formation, we can determine the enthalpy change for any reaction, which we will practice in the next section on Hess's law.

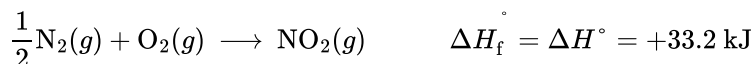
The standard enthalpy of formation of CO₂(g) is −393.5 kJ/mol. This is the enthalpy change for the exothermic reaction:

Equation:



starting with the reactants at a pressure of 1 atm and 25 °C (with the carbon present as graphite, the most stable form of carbon under these conditions) and ending with one mole of CO₂, also at 1 atm and 25 °C. For nitrogen dioxide, NO₂(g), ΔH_f° is 33.2 kJ/mol. This is the enthalpy change for the reaction:

Equation:



A reaction equation with $\frac{1}{2}$ mole of N_2 and 1 mole of O_2 is correct in this case because the standard enthalpy of formation always refers to 1 mole of product, $\text{NO}_2(g)$.

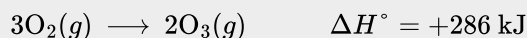
You will find a table of standard enthalpies of formation of many common substances in [Appendix G](#). These values indicate that formation reactions range from highly exothermic (such as -2984 kJ/mol for the formation of P_4O_{10}) to strongly endothermic (such as $+226.7 \text{ kJ/mol}$ for the formation of acetylene, C_2H_2). By definition, the standard enthalpy of formation of an element in its most stable form is equal to zero under standard conditions, which is 1 atm for gases and 1 M for solutions.

Example:

Evaluating an Enthalpy of Formation

Ozone, $\text{O}_3(g)$, forms from oxygen, $\text{O}_2(g)$, by an endothermic process. Ultraviolet radiation is the source of the energy that drives this reaction in the upper atmosphere. Assuming that both the reactants and products of the reaction are in their standard states, determine the standard enthalpy of formation, ΔH_f° of ozone from the following information:

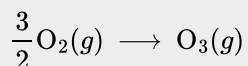
Equation:



Solution

ΔH_f° is the enthalpy change for the formation of one mole of a substance in its standard state from the elements in their standard states. Thus, ΔH_f° for $\text{O}_3(g)$ is the enthalpy change for the reaction:

Equation:



For the formation of 2 mol of $\text{O}_3(g)$, $\Delta H^\circ = +286 \text{ kJ}$. This ratio, $\left(\frac{286 \text{ kJ}}{2 \text{ mol O}_3}\right)$, can be used as a conversion factor to find the heat produced when 1 mole of $\text{O}_3(g)$ is formed, which is the enthalpy of formation for $\text{O}_3(g)$:

Equation:

$$\Delta H^\circ \text{ for 1 mole of O}_3(g) = 1 \text{ mol O}_3 \times \frac{286 \text{ kJ}}{2 \text{ mol O}_3} = 143 \text{ kJ}$$

Therefore, $\Delta H_f^\circ [\text{O}_3(g)] = +143 \text{ kJ/mol}$.

Check Your Learning

Hydrogen gas, H_2 , reacts explosively with gaseous chlorine, Cl_2 , to form hydrogen chloride, $\text{HCl}(g)$. What is the enthalpy change for the reaction of 1 mole of $\text{H}_2(g)$ with 1 mole of $\text{Cl}_2(g)$ if both the reactants and products are at standard state conditions? The standard enthalpy of formation of $\text{HCl}(g)$ is -92.3 kJ/mol .

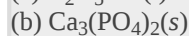
Note:

Answer:

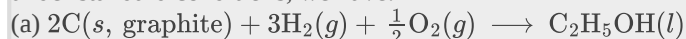
For the reaction $\text{H}_2(g) + \text{Cl}_2(g) \longrightarrow 2\text{HCl}(g) \quad \Delta H^\circ = -184.6 \text{ kJ}$

Example:**Writing Reaction Equations for ΔH_f°**

Write the heat of formation reaction equations for:

**Solution**

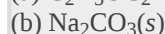
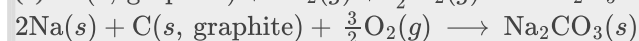
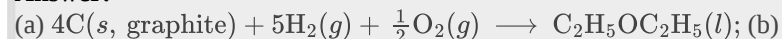
Remembering that ΔH_f° reaction equations are for forming 1 mole of the compound from its constituent elements under standard conditions, we have:



Note: The standard state of carbon is graphite, and phosphorus exists as P_4 .

Check Your Learning

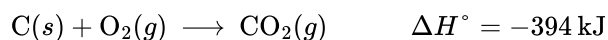
Write the heat of formation reaction equations for:

**Note:****Answer:****Hess's Law**

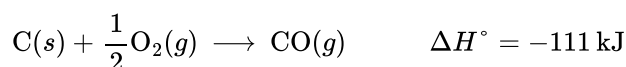
There are two ways to determine the amount of heat involved in a chemical change: measure it experimentally, or calculate it from other experimentally determined enthalpy changes. Some reactions are difficult, if not impossible, to investigate and make accurate measurements for experimentally. And even when a reaction is not hard to perform or measure, it is convenient to be able to determine the heat involved in a reaction without having to perform an experiment.

This type of calculation usually involves the use of **Hess's law**, which states: *If a process can be written as the sum of several stepwise processes, the enthalpy change of the total process equals the sum of the enthalpy changes of the various steps.* Hess's law is valid because enthalpy is a state function: Enthalpy changes depend only on where a chemical process starts and ends, but not on the path it takes from start to finish. For example, we can think of the reaction of carbon with oxygen to form carbon dioxide as occurring either directly or by a two-step process.

The direct process is written:

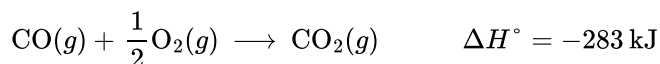
Equation:

In the two-step process, first carbon monoxide is formed:

Equation:

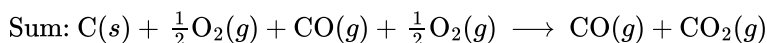
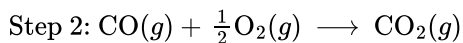
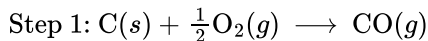
Then, carbon monoxide reacts further to form carbon dioxide:

Equation:



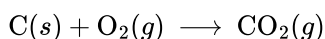
The equation describing the overall reaction is the sum of these two chemical changes:

Equation:



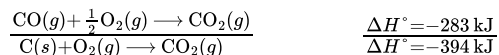
Because the CO produced in Step 1 is consumed in Step 2, the net change is:

Equation:

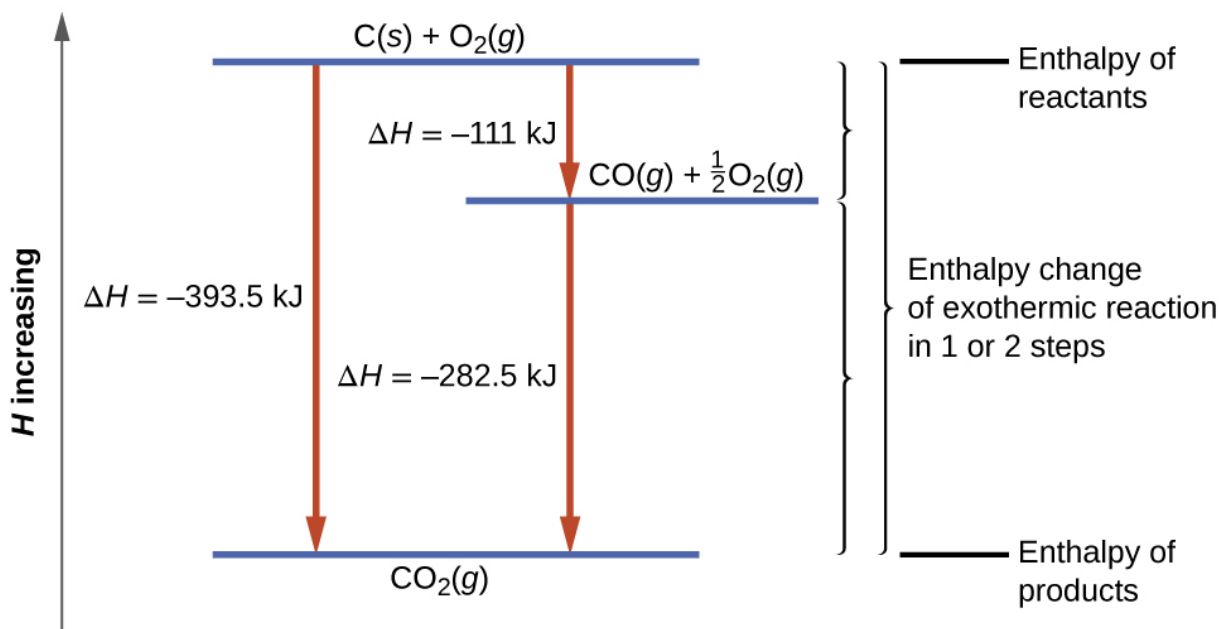


According to Hess's law, the enthalpy change of the reaction will equal the sum of the enthalpy changes of the steps.

Equation:



The result is shown in [\[link\]](#). We see that ΔH of the overall reaction is the same whether it occurs in one step or two. This finding (overall ΔH for the reaction = sum of ΔH values for reaction "steps" in the overall reaction) is true in general for chemical and physical processes.

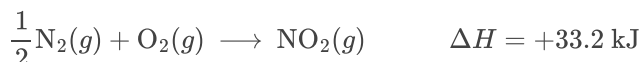


The formation of $\text{CO}_2(g)$ from its elements can be thought of as occurring in two steps, which sum to the overall reaction, as described by Hess's law. The horizontal blue lines represent enthalpies. For an exothermic process, the products are at lower enthalpy than are the reactants.

Before we further practice using Hess's law, let us recall two important features of ΔH .

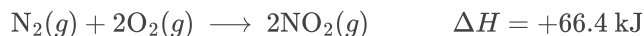
1. ΔH is directly proportional to the quantities of reactants or products. For example, the enthalpy change for the reaction forming 1 mole of $\text{NO}_2(g)$ is +33.2 kJ:

Equation:



When 2 moles of NO_2 (twice as much) are formed, the ΔH will be twice as large:

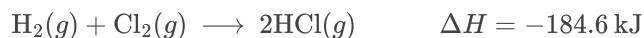
Equation:



In general, if we multiply or divide an equation by a number, then the enthalpy change should also be multiplied or divided by the same number.

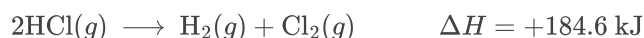
2. ΔH for a reaction in one direction is equal in magnitude and opposite in sign to ΔH for the reaction in the reverse direction. For example, given that:

Equation:



Then, for the "reverse" reaction, the enthalpy change is also "reversed":

Equation:

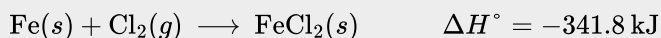


Example:

Stepwise Calculation of ΔH_f° Using Hess's Law

Determine the enthalpy of formation, ΔH_f° , of $\text{FeCl}_3(s)$ from the enthalpy changes of the following two-step process that occurs under standard state conditions:

Equation:



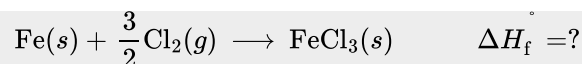
Equation:



Solution

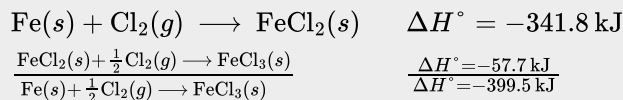
We are trying to find the standard enthalpy of formation of $\text{FeCl}_3(s)$, which is equal to ΔH° for the reaction:

Equation:



Looking at the reactions, we see that the reaction for which we want to find ΔH° is the sum of the two reactions with known ΔH values, so we must sum their ΔH s:

Equation:

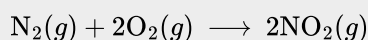


The enthalpy of formation, ΔH_f° , of $\text{FeCl}_3(s)$ is -399.5 kJ/mol .

Check Your Learning

Calculate ΔH for the process:

Equation:

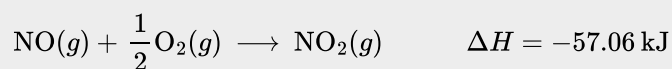


from the following information:

Equation:



Equation:



Note:

Answer:

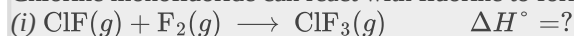
66.4 kJ

Here is a less straightforward example that illustrates the thought process involved in solving many Hess's law problems. It shows how we can find many standard enthalpies of formation (and other values of ΔH) if they are difficult to determine experimentally.

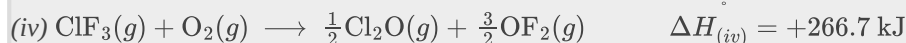
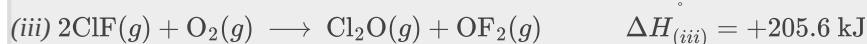
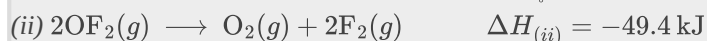
Example:

A More Challenging Problem Using Hess's Law

Chlorine monofluoride can react with fluorine to form chlorine trifluoride:



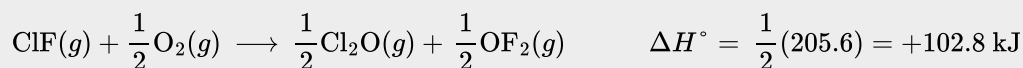
Use the reactions here to determine the ΔH° for reaction (i):



Solution

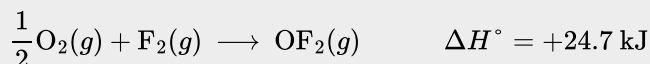
Our goal is to manipulate and combine reactions (ii), (iii), and (iv) such that they add up to reaction (i). Going from left to right in (i), we first see that ClF(g) is needed as a reactant. This can be obtained by multiplying reaction (iii) by $\frac{1}{2}$, which means that the ΔH° change is also multiplied by $\frac{1}{2}$:

Equation:



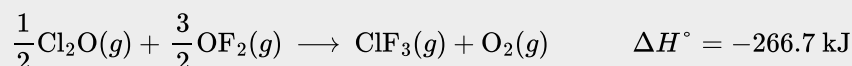
Next, we see that F₂ is also needed as a reactant. To get this, reverse and halve reaction (ii), which means that the ΔH° changes sign and is halved:

Equation:



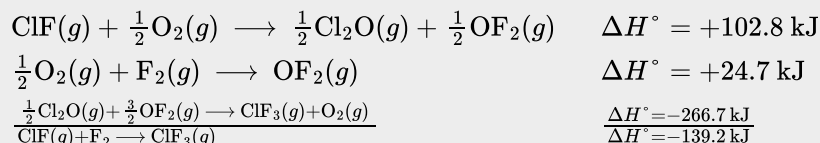
To get ClF₃ as a product, reverse (iv), changing the sign of ΔH° :

Equation:



Now check to make sure that these reactions add up to the reaction we want:

Equation:



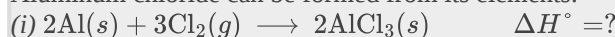
Reactants $\frac{1}{2}\text{O}_2$ and $\frac{1}{2}\text{O}_2$ cancel out product O₂; product $\frac{1}{2}\text{Cl}_2\text{O}$ cancels reactant $\frac{1}{2}\text{Cl}_2\text{O}$; and reactant $\frac{3}{2}\text{OF}_2$ is cancelled by products $\frac{1}{2}\text{OF}_2$ and OF₂. This leaves only reactants ClF(g) and F₂(g) and product ClF₃(g), which are what we want. Since summing these three modified reactions yields the reaction of interest, summing the three modified ΔH° values will give the desired ΔH° :

Equation:

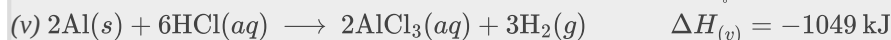
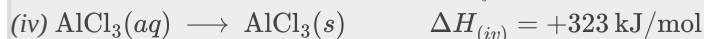
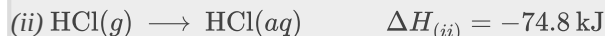
$$\Delta H^\circ = (+102.8 \text{ kJ}) + (24.7 \text{ kJ}) + (-266.7 \text{ kJ}) = -139.2 \text{ kJ}$$

Check Your Learning

Aluminum chloride can be formed from its elements:



Use the reactions here to determine the ΔH° for reaction (i):



Note:

Answer:

-1407 kJ

We also can use Hess's law to determine the enthalpy change of any reaction if the corresponding enthalpies of formation of the reactants and products are available. The stepwise reactions we consider are: (i) decompositions of the reactants into their component elements (for which the enthalpy changes are proportional to the negative of the enthalpies of formation of the reactants), followed by (ii) re-combinations of the elements to give the products (with the enthalpy changes proportional to the enthalpies of formation of the products). The standard enthalpy change of the overall reaction is therefore equal to: (ii) the sum of the standard enthalpies of formation of all the products plus (i) the sum of the negatives of the standard enthalpies of formation of the reactants. This is usually rearranged slightly to be written as follows, with Σ representing "the sum of" and n standing for the stoichiometric coefficients:

Equation:

$$\Delta H_{\text{reaction}} = \sum n \times \Delta H_f^\circ (\text{products}) - \sum n \times \Delta H_f^\circ (\text{reactants})$$

The following example shows in detail why this equation is valid, and how to use it to calculate the enthalpy change for a reaction of interest.

Example:

Using Hess's Law

What is the standard enthalpy change for the reaction:

Equation:



Solution: Using the Equation

Use the special form of Hess's law given previously, and values from [Appendix G](#):

Equation:

$$\Delta H_{\text{reaction}} = \sum n \times \Delta H_f^\circ (\text{products}) - \sum n \times \Delta H_f^\circ (\text{reactants})$$

Equation:

$$\begin{aligned} &= \left[2 \cancel{\text{mol HNO}_3(aq)} \times \frac{-207.4 \text{ kJ}}{\cancel{\text{mol HNO}_3(aq)}} + 1 \cancel{\text{mol NO}(g)} \times \frac{+90.2 \text{ kJ}}{\cancel{\text{mol NO}(g)}} \right] \\ &- \left[3 \cancel{\text{mol NO}_2(g)} \times \frac{+33.2 \text{ kJ}}{\cancel{\text{mol NO}_2(g)}} + 1 \cancel{\text{mol H}_2\text{O}(l)} \times \frac{-285.8 \text{ kJ}}{\cancel{\text{mol H}_2\text{O}(l)}} \right] \\ &= [2 \times (-206.64) + 90.25] - [3 \times 33.2 + -(-285.83)] \\ &= -323.03 + 186.23 \\ &= -136.80 \text{ kJ} \end{aligned}$$

Solution: Supporting Why the General Equation Is Valid

Alternatively, we can write this reaction as the sum of the decompositions of $3\text{NO}_2(g)$ and $1\text{H}_2\text{O}(l)$ into their constituent elements, and the formation of $2\text{HNO}_3(aq)$ and $1\text{NO}(g)$ from their constituent elements. Writing out these reactions, and noting their relationships to the ΔH_f° values for these compounds (from [Appendix G](#)), we have:

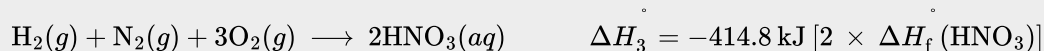
Equation:



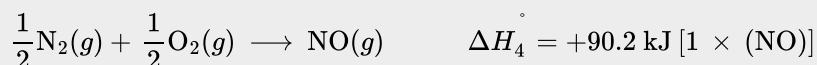
Equation:



Equation:

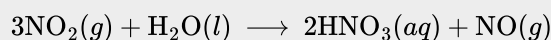


Equation:



Summing these reaction equations gives the reaction we are interested in:

Equation:



Summing their enthalpy changes gives the value we want to determine:

Equation:

$$\begin{aligned} \Delta H_{\text{rxn}}^\circ &= \Delta H_1^\circ + \Delta H_2^\circ + \Delta H_3^\circ + \Delta H_4^\circ = (-99.6 \text{ kJ}) + (+285.8 \text{ kJ}) + (-414.8 \text{ kJ}) + (+90.2 \text{ kJ}) \\ &= -138.4 \text{ kJ} \end{aligned}$$

So the standard enthalpy change for this reaction is $\Delta H^\circ = -138.4 \text{ kJ}$.

Note that this result was obtained by (1) multiplying the ΔH_f° of each product by its stoichiometric coefficient and summing those values, (2) multiplying the ΔH_f° of each reactant by its stoichiometric coefficient and summing those values, and then (3) subtracting the result found in (2) from the result found in (1). This is also the procedure in using the general equation, as shown.

Check Your Learning

Calculate the heat of combustion of 1 mole of ethanol, $\text{C}_2\text{H}_5\text{OH}(l)$, when $\text{H}_2\text{O}(l)$ and $\text{CO}_2(g)$ are formed. Use the following enthalpies of formation: $\text{C}_2\text{H}_5\text{OH}(l)$, -278 kJ/mol ; $\text{H}_2\text{O}(l)$, -286 kJ/mol ; and $\text{CO}_2(g)$, -394 kJ/mol .

Note:

Answer:

-1368 kJ/mol

Key Concepts and Summary

If a chemical change is carried out at constant pressure and the only work done is caused by expansion or contraction, q for the change is called the enthalpy change with the symbol ΔH , or ΔH° for reactions occurring under standard state conditions at 298 K. The value of ΔH for a reaction in one direction is equal in magnitude, but opposite in sign, to ΔH for the reaction in the opposite direction, and ΔH is directly proportional to the quantity of reactants and products. The standard enthalpy of formation, ΔH_f° , is the enthalpy change accompanying the formation of 1 mole of a substance from the elements in their most stable states at 1 bar and 298.15 K. If the enthalpies of formation are available for the reactants and products of a reaction, the enthalpy change can be calculated using Hess's law: If a process can be written as the sum of several stepwise processes, the enthalpy change of the total process equals the sum of the enthalpy changes of the various steps.

Key Equations

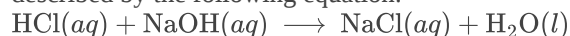
- $\Delta U = q + w$
- $\Delta H_{\text{reaction}} = \sum n \times \Delta H_{\text{f}}^{\circ}(\text{products}) - \sum n \times \Delta H_{\text{f}}^{\circ}(\text{reactants})$

Chemistry End of Chapter Exercises

Exercise:

Problem:

Explain how the heat measured in [\[link\]](#) differs from the enthalpy change for the exothermic reaction described by the following equation:



Solution:

The enthalpy change of the indicated reaction is for exactly 1 mol HCL and 1 mol NaOH; the heat in the example is produced by 0.0500 mol HCl and 0.0500 mol NaOH.

Exercise:

Problem:

Using the data in the check your learning section of [\[link\]](#), calculate ΔH in kJ/mol of $\text{AgNO}_3(aq)$ for the reaction: $\text{NaCl}(aq) + \text{AgNO}_3(aq) \longrightarrow \text{AgCl}(s) + \text{NaNO}_3(aq)$

Exercise:

Problem:

Calculate the enthalpy of solution (ΔH for the dissolution) per mole of NH_4NO_3 under the conditions described in [\[link\]](#).

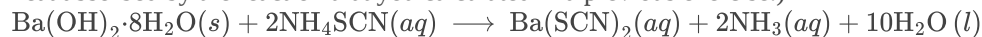
Solution:

25 kJ mol⁻¹

Exercise:

Problem:

Calculate ΔH for the reaction described by the equation. (*Hint:* Use the value for the approximate amount of heat absorbed by the reaction that you calculated in a previous exercise.)



Exercise:

Problem: Calculate the enthalpy of solution (ΔH for the dissolution) per mole of CaCl_2 (refer to [\[link\]](#)).

Solution:

81 kJ mol⁻¹

Exercise:

Problem:

Although the gas used in an oxyacetylene torch ([link](#)) is essentially pure acetylene, the heat produced by combustion of one mole of acetylene in such a torch is likely not equal to the enthalpy of combustion of acetylene listed in [link](#). Considering the conditions for which the tabulated data are reported, suggest an explanation.

Exercise:

Problem: How much heat is produced by burning 4.00 moles of acetylene under standard state conditions?

Solution:

5204.4 kJ

Exercise:

Problem: How much heat is produced by combustion of 125 g of methanol under standard state conditions?

Exercise:**Problem:**

How many moles of isooctane must be burned to produce 100 kJ of heat under standard state conditions?

Solution:

1.83×10^{-2} mol

Exercise:**Problem:**

What mass of carbon monoxide must be burned to produce 175 kJ of heat under standard state conditions?

Exercise:**Problem:**

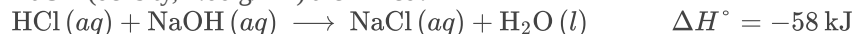
When 2.50 g of methane burns in oxygen, 125 kJ of heat is produced. What is the enthalpy of combustion per mole of methane under these conditions?

Solution:

-802 kJ mol^{-1}

Exercise:**Problem:**

How much heat is produced when 100 mL of 0.250 M HCl (density, 1.00 g/mL) and 200 mL of 0.150 M NaOH (density, 1.00 g/mL) are mixed?



If both solutions are at the same temperature and the heat capacity of the products is 4.19 J/g °C, how much will the temperature increase? What assumption did you make in your calculation?

Exercise:

Problem:

A sample of 0.562 g of carbon is burned in oxygen in a bomb calorimeter, producing carbon dioxide. Assume both the reactants and products are under standard state conditions, and that the heat released is directly proportional to the enthalpy of combustion of graphite. The temperature of the calorimeter increases from 26.74 °C to 27.93 °C. What is the heat capacity of the calorimeter and its contents?

Solution:

15.5 kJ/°C

Exercise:**Problem:**

Before the introduction of chlorofluorocarbons, sulfur dioxide (enthalpy of vaporization, 6.00 kcal/mol) was used in household refrigerators. What mass of SO₂ must be evaporated to remove as much heat as evaporation of 1.00 kg of CCl₂F₂ (enthalpy of vaporization is 17.4 kJ/mol)?

The vaporization reactions for SO₂ and CCl₂F₂ are $\text{SO}_2(l) \longrightarrow \text{SO}_2(g)$ and $\text{CCl}_2\text{F}_2(l) \longrightarrow \text{CCl}_2\text{F}_2(g)$, respectively.

Exercise:**Problem:**

Homes may be heated by pumping hot water through radiators. What mass of water will provide the same amount of heat when cooled from 95.0 to 35.0 °C, as the heat provided when 100 g of steam is cooled from 110 °C to 100 °C.

Solution:

7.43 g

Exercise:

Problem: Which of the enthalpies of combustion in [\[link\]](#) the table are also standard enthalpies of formation?

Exercise:**Problem:**

Does the standard enthalpy of formation of H₂O(g) differ from ΔH° for the reaction $2\text{H}_2(g) + \text{O}_2(g) \longrightarrow 2\text{H}_2\text{O}(g)$?

Solution:

Yes.

Exercise:**Problem:**

Joseph Priestly prepared oxygen in 1774 by heating red mercury(II) oxide with sunlight focused through a lens. How much heat is required to decompose exactly 1 mole of red HgO(s) to Hg(l) and O₂(g) under standard conditions?

Exercise:**Problem:**

How many kilojoules of heat will be released when exactly 1 mole of manganese, Mn, is burned to form Mn₃O₄(s) at standard state conditions?

Solution:

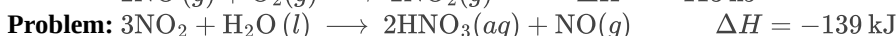
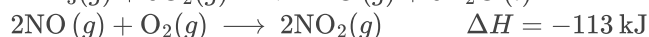
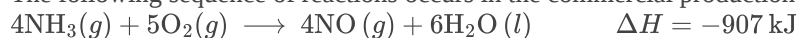
459.6 kJ

Exercise:**Problem:**

How many kilojoules of heat will be released when exactly 1 mole of iron, Fe, is burned to form $\text{Fe}_2\text{O}_3(\text{s})$ at standard state conditions?

Exercise:

The following sequence of reactions occurs in the commercial production of aqueous nitric acid:



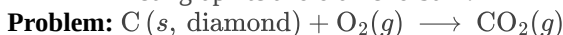
Determine the total energy change for the production of one mole of aqueous nitric acid by this process.

Solution:

-495 kJ/mol

Exercise:

Both graphite and diamond burn.



For the conversion of graphite to diamond:



Which produces more heat, the combustion of graphite or the combustion of diamond?

Exercise:**Problem:**

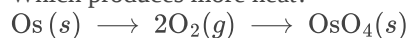
From the molar heats of formation in [Appendix G](#), determine how much heat is required to evaporate one mole of water: $\text{H}_2\text{O}(\text{l}) \longrightarrow \text{H}_2\text{O}(\text{g})$

Solution:

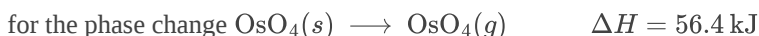
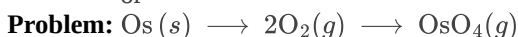
44.01 kJ/mol

Exercise:

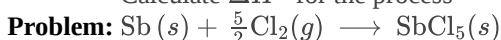
Which produces more heat?



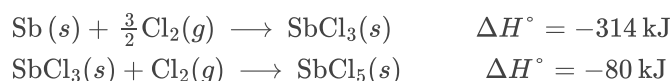
or

**Exercise:**

Calculate ΔH° for the process



from the following information:



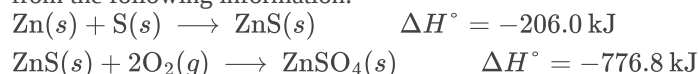
Solution:

-394 kJ

Exercise:

Problem: Calculate ΔH° for the process $\text{Zn}(s) + \text{S}(s) + 2\text{O}_2(g) \longrightarrow \text{ZnSO}_4(s)$

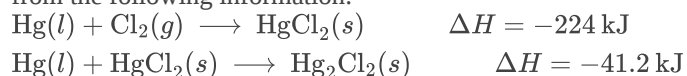
from the following information:



Exercise:

Problem: Calculate ΔH for the process $\text{Hg}_2\text{Cl}_2(s) \longrightarrow 2\text{Hg}(l) + \text{Cl}_2(g)$

from the following information:



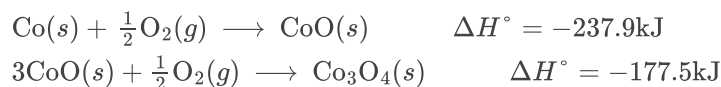
Solution:

265 kJ

Exercise:

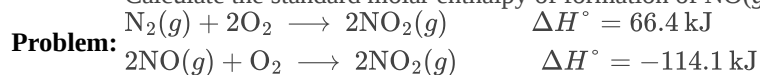
Problem: Calculate ΔH° for the process $\text{Co}_3\text{O}_4(s) \longrightarrow 3\text{Co}(s) + 2\text{O}_2(g)$

from the following information:



Exercise:

Calculate the standard molar enthalpy of formation of $\text{NO}(g)$ from the following data:



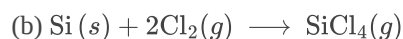
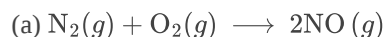
Solution:

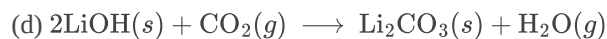
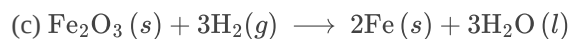
90.3 kJ/mol

Exercise:

Problem:

Using the data in [Appendix G](#), calculate the standard enthalpy change for each of the following reactions:

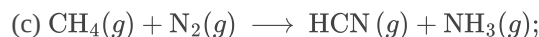
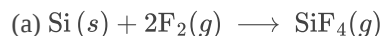




Exercise:

Problem:

Using the data in [Appendix G](#), calculate the standard enthalpy change for each of the following reactions:



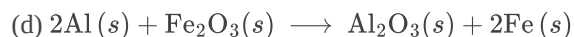
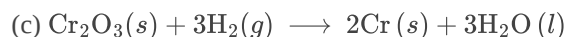
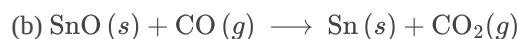
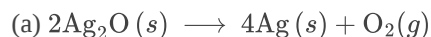
Solution:

(a) $-1615.0 \text{ kJ mol}^{-1}$; (b) $-484.3 \text{ kJ mol}^{-1}$; (c) 164.2 kJ ; (d) -232.1 kJ

Exercise:

Problem:

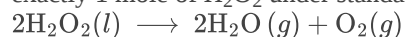
The following reactions can be used to prepare samples of metals. Determine the enthalpy change under standard state conditions for each.



Exercise:

Problem:

The decomposition of hydrogen peroxide, H_2O_2 , has been used to provide thrust in the control jets of various space vehicles. Using the data in [Appendix G](#), determine how much heat is produced by the decomposition of exactly 1 mole of H_2O_2 under standard conditions.



Solution:

$-54.04 \text{ kJ mol}^{-1}$

Exercise:

Problem:

Calculate the enthalpy of combustion of propane, $\text{C}_3\text{H}_8(g)$, for the formation of $\text{H}_2\text{O}(g)$ and $\text{CO}_2(g)$. The enthalpy of formation of propane is -104 kJ/mol .

Exercise:

Problem:

Calculate the enthalpy of combustion of butane, $\text{C}_4\text{H}_{10}(g)$ for the formation of $\text{H}_2\text{O}(g)$ and $\text{CO}_2(g)$. The enthalpy of formation of butane is -126 kJ/mol .

Solution:

$-2660 \text{ kJ mol}^{-1}$

Exercise:**Problem:**

Both propane and butane are used as gaseous fuels. Which compound produces more heat per gram when burned?

Exercise:**Problem:**

The white pigment TiO_2 is prepared by the reaction of titanium tetrachloride, TiCl_4 , with water vapor in the gas phase: $\text{TiCl}_4(g) + 2\text{H}_2\text{O}(g) \longrightarrow \text{TiO}_2(s) + 4\text{HCl}(g)$.

How much heat is evolved in the production of exactly 1 mole of $\text{TiO}_2(s)$ under standard state conditions?

Solution:

67.1 kJ

Exercise:**Problem:**

Water gas, a mixture of H_2 and CO , is an important industrial fuel produced by the reaction of steam with red hot coke, essentially pure carbon: $\text{C}(s) + \text{H}_2\text{O}(g) \longrightarrow \text{CO}(g) + \text{H}_2(g)$.

(a) Assuming that coke has the same enthalpy of formation as graphite, calculate ΔH° for this reaction.

(b) Methanol, a liquid fuel that could possibly replace gasoline, can be prepared from water gas and additional hydrogen at high temperature and pressure in the presence of a suitable catalyst:
 $2\text{H}_2(g) + \text{CO}(g) \longrightarrow \text{CH}_3\text{OH}(g)$.

Under the conditions of the reaction, methanol forms as a gas. Calculate ΔH° for this reaction and for the condensation of gaseous methanol to liquid methanol.

(c) Calculate the heat of combustion of 1 mole of liquid methanol to $\text{H}_2\text{O}(g)$ and $\text{CO}_2(g)$.

Exercise:**Problem:**

In the early days of automobiles, illumination at night was provided by burning acetylene, C_2H_2 . Though no longer used as auto headlamps, acetylene is still used as a source of light by some cave explorers. The acetylene is (was) prepared in the lamp by the reaction of water with calcium carbide, CaC_2 :
 $\text{CaC}_2(s) + 2\text{H}_2\text{O}(l) \longrightarrow \text{Ca}(\text{OH})_2(s) + \text{C}_2\text{H}_2(g)$.

Calculate the standard enthalpy of the reaction. The ΔH_f° of CaC_2 is -15.14 kcal/mol .

Solution:

-122.8 kJ

Exercise:**Problem:**

From the data in [\[link\]](#), determine which of the following fuels produces the greatest amount of heat per gram when burned under standard conditions: $\text{CO}(g)$, $\text{CH}_4(g)$, or $\text{C}_2\text{H}_2(g)$.

Exercise:**Problem:**

The enthalpy of combustion of hard coal averages -35 kJ/g , that of gasoline, $1.28 \times 10^5 \text{ kJ/gal}$. How many kilograms of hard coal provide the same amount of heat as is available from 1.0 gallon of gasoline? Assume that the density of gasoline is 0.692 g/mL (the same as the density of isooctane).

Solution:

3.7 kg

Exercise:

Problem: Ethanol, $\text{C}_2\text{H}_5\text{OH}$, is used as a fuel for motor vehicles, particularly in Brazil.

(a) Write the balanced equation for the combustion of ethanol to $\text{CO}_2(g)$ and $\text{H}_2\text{O}(g)$, and, using the data in [Appendix G](#), calculate the enthalpy of combustion of 1 mole of ethanol.

(b) The density of ethanol is 0.7893 g/mL . Calculate the enthalpy of combustion of exactly 1 L of ethanol.

(c) Assuming that an automobile's mileage is directly proportional to the heat of combustion of the fuel, calculate how much farther an automobile could be expected to travel on 1 L of gasoline than on 1 L of ethanol. Assume that gasoline has the heat of combustion and the density of n-octane, C_8H_{18}

($\Delta H_f^\circ = -208.4 \text{ kJ/mol}$; density = 0.7025 g/mL).

Exercise:**Problem:**

Among the substances that react with oxygen and that have been considered as potential rocket fuels are diborane [B_2H_6 , produces $\text{B}_2\text{O}_3(s)$ and $\text{H}_2\text{O}(g)$], methane [CH_4 , produces $\text{CO}_2(g)$ and $\text{H}_2\text{O}(g)$], and hydrazine [N_2H_4 , produces $\text{N}_2(g)$ and $\text{H}_2\text{O}(g)$]. On the basis of the heat released by 1.00 g of each substance in its reaction with oxygen, which of these compounds offers the best possibility as a rocket fuel? The ΔH_f° of $\text{B}_2\text{H}_6(g)$, $\text{CH}_4(g)$, and $\text{N}_2\text{H}_4(l)$ may be found in [Appendix G](#).

Solution:

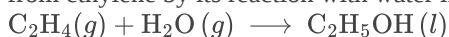
On the assumption that the best rocket fuel is the one that gives off the most heat, B_2H_6 is the prime candidate.

Exercise:**Problem:**

How much heat is produced when 1.25 g of chromium metal reacts with oxygen gas under standard conditions?

Exercise:**Problem:**

Ethylene, C_2H_2 , a byproduct from the fractional distillation of petroleum, is fourth among the 50 chemical compounds produced commercially in the largest quantities. About 80% of synthetic ethanol is manufactured from ethylene by its reaction with water in the presence of a suitable catalyst.



Using the data in the table in [Appendix G](#), calculate ΔH° for the reaction.

Solution:

-88.2 kJ

Exercise:

The oxidation of the sugar glucose, $\text{C}_6\text{H}_{12}\text{O}_6$, is described by the following equation:



The metabolism of glucose gives the same products, although the glucose reacts with oxygen in a series of steps in the body.

(a) How much heat in kilojoules can be produced by the metabolism of 1.0 g of glucose?

(b) How many Calories can be produced by the metabolism of 1.0 g of glucose?

Exercise:

Problem: Propane, C_3H_8 , is a hydrocarbon that is commonly used as a fuel.

(a) Write a balanced equation for the complete combustion of propane gas.

(b) Calculate the volume of air at 25 °C and 1.00 atmosphere that is needed to completely combust 25.0 grams of propane. Assume that air is 21.0 percent O_2 by volume. (Hint: We will see how to do this calculation in a later chapter on gases—for now use the information that 1.00 L of air at 25 °C and 1.00 atm contains 0.275 g of O_2 per liter.)

(c) The heat of combustion of propane is -2,219.2 kJ/mol. Calculate the heat of formation, ΔH_f° of propane given that ΔH_f° of $\text{H}_2\text{O}(l) = -285.8 \text{ kJ/mol}$ and ΔH_f° of $\text{CO}_2(g) = -393.5 \text{ kJ/mol}$.

(d) Assuming that all of the heat released in burning 25.0 grams of propane is transferred to 4.00 kilograms of water, calculate the increase in temperature of the water.

Solution:

(a) $\text{C}_3\text{H}_8(g) + 5\text{O}_2(g) \longrightarrow 3\text{CO}_2(g) + 4\text{H}_2\text{O}(l)$; (b) 330 L; (c) $-104.5 \text{ kJ mol}^{-1}$; (d) 75.4 °C

Exercise:

Problem:

During a recent winter month in Sheboygan, Wisconsin, it was necessary to obtain 3500 kWh of heat provided by a natural gas furnace with 89% efficiency to keep a small house warm (the efficiency of a gas furnace is the percent of the heat produced by combustion that is transferred into the house).

(a) Assume that natural gas is pure methane and determine the volume of natural gas in cubic feet that was required to heat the house. The average temperature of the natural gas was 56 °F; at this temperature and a pressure of 1 atm, natural gas has a density of 0.681 g/L.

(b) How many gallons of LPG (liquefied petroleum gas) would be required to replace the natural gas used? Assume the LPG is liquid propane [C_3H_8 ; density, 0.5318 g/mL; enthalpy of combustion, 2219 kJ/mol for the formation of $\text{CO}_2(g)$ and $\text{H}_2\text{O}(l)$] and the furnace used to burn the LPG has the same efficiency as the gas furnace.

(c) What mass of carbon dioxide is produced by combustion of the methane used to heat the house?

(d) What mass of water is produced by combustion of the methane used to heat the house?

(e) What volume of air is required to provide the oxygen for the combustion of the methane used to heat the house? Air contains 23% oxygen by mass. The average density of air during the month was 1.22 g/L.

(f) How many kilowatt-hours ($1 \text{ kWh} = 3.6 \times 10^6 \text{ J}$) of electricity would be required to provide the heat necessary to heat the house? Note electricity is 100% efficient in producing heat inside a house.

(g) Although electricity is 100% efficient in producing heat inside a house, production and distribution of electricity is not 100% efficient. The efficiency of production and distribution of electricity produced in a coal-fired power plant is about 40%. A certain type of coal provides 2.26 kWh per pound upon combustion. What mass of this coal in kilograms will be required to produce the electrical energy necessary to heat the house if the efficiency of generation and distribution is 40%?

Glossary

chemical thermodynamics

area of science that deals with the relationships between heat, work, and all forms of energy associated with chemical and physical processes

enthalpy (H)

sum of a system's internal energy and the mathematical product of its pressure and volume

enthalpy change (ΔH)

heat released or absorbed by a system under constant pressure during a chemical or physical process

expansion work (pressure-volume work)

work done as a system expands or contracts against external pressure

first law of thermodynamics

internal energy of a system changes due to heat flow in or out of the system or work done on or by the system

Hess's law

if a process can be represented as the sum of several steps, the enthalpy change of the process equals the sum of the enthalpy changes of the steps

hydrocarbon

compound composed only of hydrogen and carbon; the major component of fossil fuels

internal energy (U)

total of all possible kinds of energy present in a substance or substances

standard enthalpy of combustion (ΔH_c°)

heat released when one mole of a compound undergoes complete combustion under standard conditions

standard enthalpy of formation (ΔH_f°)

enthalpy change of a chemical reaction in which 1 mole of a pure substance is formed from its elements in their most stable states under standard state conditions

standard state

set of physical conditions as accepted as common reference conditions for reporting thermodynamic properties; 1 bar of pressure, and solutions at 1 molar concentrations, usually at a temperature of 298.15 K

state function

property depending only on the state of a system, and not the path taken to reach that state

Electromagnetic Energy

By the end of this section, you will be able to:

- Explain the basic behavior of waves, including travelling waves and standing waves
- Describe the wave nature of light
- Use appropriate equations to calculate related light-wave properties such as frequency, wavelength, and energy
- Distinguish between line and continuous emission spectra
- Describe the particle nature of light

The nature of light has been a subject of inquiry since antiquity. In the seventeenth century, Isaac Newton performed experiments with lenses and prisms and was able to demonstrate that white light consists of the individual colors of the rainbow combined together. Newton explained his optics findings in terms of a "corpuscular" view of light, in which light was composed of streams of extremely tiny particles travelling at high speeds according to Newton's laws of motion. Others in the seventeenth century, such as Christiaan Huygens, had shown that optical phenomena such as reflection and refraction could be equally well explained in terms of light as waves travelling at high speed through a medium called "luminiferous aether" that was thought to permeate all space. Early in the nineteenth century, Thomas Young demonstrated that light passing through narrow, closely spaced slits produced interference patterns that could not be explained in terms of Newtonian particles but could be easily explained in terms of waves. Later in the nineteenth century, after James Clerk Maxwell developed his theory of **electromagnetic radiation** and showed that light was the visible part of a vast spectrum of electromagnetic waves, the particle view of light became thoroughly discredited. By the end of the nineteenth century, scientists viewed the physical universe as roughly comprising two separate domains: matter composed of particles moving according to Newton's laws of motion, and electromagnetic radiation consisting of waves governed by Maxwell's equations. Today, these domains are referred to as classical mechanics and classical electrodynamics (or classical electromagnetism). Although there were a few physical phenomena that could not be explained within this framework, scientists at that time were so confident of the overall soundness of this

framework that they viewed these aberrations as puzzling paradoxes that would ultimately be resolved somehow within this framework. As we shall see, these paradoxes led to a contemporary framework that intimately connects particles and waves at a fundamental level called wave-particle duality, which has superseded the classical view.

Visible light and other forms of electromagnetic radiation play important roles in chemistry, since they can be used to infer the energies of electrons within atoms and molecules. Much of modern technology is based on electromagnetic radiation. For example, radio waves from a mobile phone, X-rays used by dentists, the energy used to cook food in your microwave, the radiant heat from red-hot objects, and the light from your television screen are forms of electromagnetic radiation that all exhibit wavelike behavior.

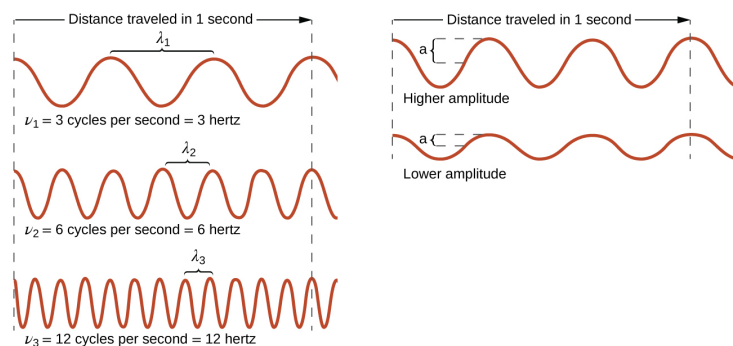
Waves

A **wave** is an oscillation or periodic movement that can transport energy from one point in space to another. Common examples of waves are all around us. Shaking the end of a rope transfers energy from your hand to the other end of the rope, dropping a pebble into a pond causes waves to ripple outward along the water's surface, and the expansion of air that accompanies a lightning strike generates sound waves (thunder) that can travel outward for several miles. In each of these cases, kinetic energy is transferred through matter (the rope, water, or air) while the matter remains essentially in place. An insightful example of a wave occurs in sports stadiums when fans in a narrow region of seats rise simultaneously and stand with their arms raised up for a few seconds before sitting down again while the fans in neighboring sections likewise stand up and sit down in sequence. While this wave can quickly encircle a large stadium in a few seconds, none of the fans actually travel with the wave—they all stay in or above their seats.

Waves need not be restricted to travel through matter. As Maxwell showed, electromagnetic waves consist of an electric field oscillating in step with a perpendicular magnetic field, both of which are perpendicular to the

direction of travel. These waves can travel through a vacuum at a constant speed of 2.998×10^8 m/s, the speed of light (denoted by c).

All waves, including forms of electromagnetic radiation, are characterized by, a **wavelength** (denoted by λ , the lowercase Greek letter lambda), a **frequency** (denoted by ν , the lowercase Greek letter nu), and an **amplitude**. As can be seen in [\[link\]](#), the wavelength is the distance between two consecutive peaks or troughs in a wave (measured in meters in the SI system). Electromagnetic waves have wavelengths that fall within an enormous range-wavelengths of kilometers (10^3 m) to picometers (10^{-12} m) have been observed. The frequency is the number of wave cycles that pass a specified point in space in a specified amount of time (in the SI system, this is measured in seconds). A cycle corresponds to one complete wavelength. The unit for frequency, expressed as cycles per second [s^{-1}], is the **hertz (Hz)**. Common multiples of this unit are megahertz, ($1 \text{ MHz} = 1 \times 10^6 \text{ Hz}$) and gigahertz ($1 \text{ GHz} = 1 \times 10^9 \text{ Hz}$). The amplitude corresponds to the magnitude of the wave's displacement and so, in [\[link\]](#), this corresponds to one-half the height between the peaks and troughs. The amplitude is related to the intensity of the wave, which for light is the brightness, and for sound is the loudness.



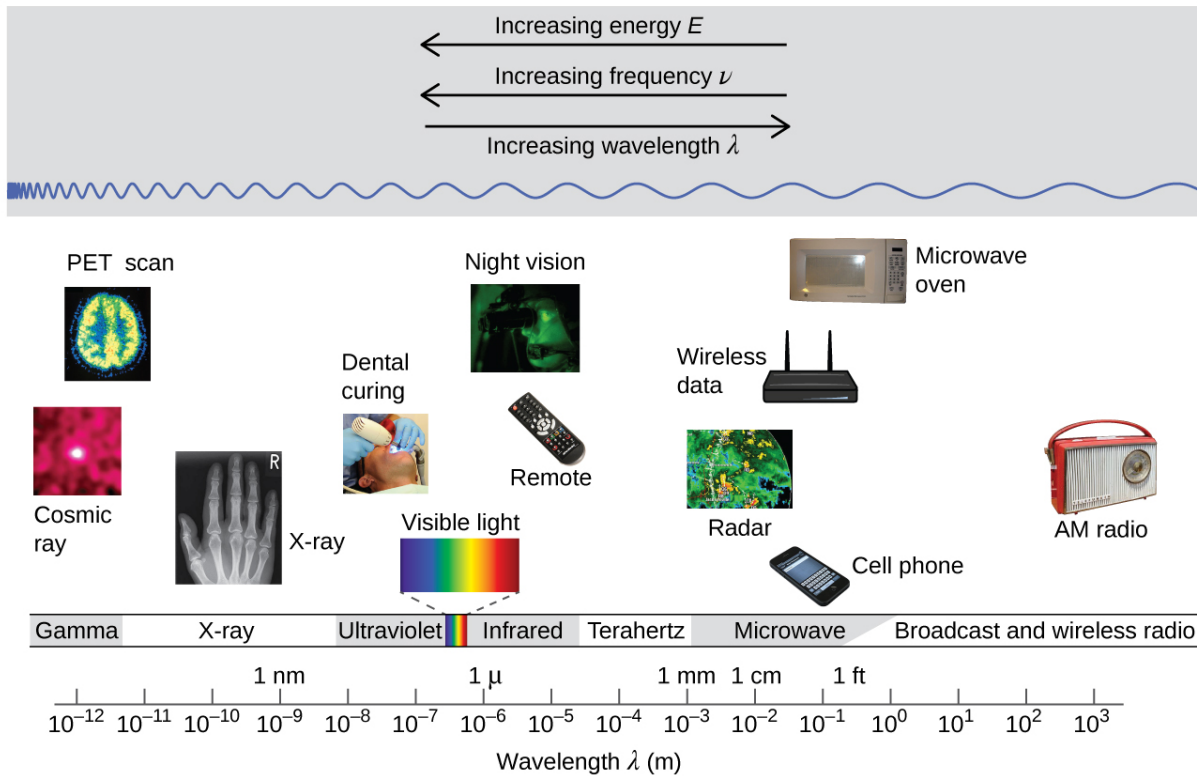
One-dimensional sinusoidal waves show the relationship among wavelength, frequency, and speed. The wave with the shortest wavelength has the highest frequency. Amplitude is one-half the height of the wave from peak to trough.

The product of a wave's wavelength (λ) and its frequency (ν), $\lambda\nu$, is the speed of the wave. Thus, for electromagnetic radiation in a vacuum, speed is equal to the fundamental constant, c :

Equation:

$$c = 2.998 \times 10^8 \text{ms}^{-1} = \lambda\nu$$

Wavelength and frequency are inversely proportional: As the wavelength increases, the frequency decreases. The inverse proportionality is illustrated in [\[link\]](#). This figure also shows the **electromagnetic spectrum**, the range of all types of electromagnetic radiation. Each of the various colors of visible light has specific frequencies and wavelengths associated with them, and you can see that visible light makes up only a small portion of the electromagnetic spectrum. Because the technologies developed to work in various parts of the electromagnetic spectrum are different, for reasons of convenience and historical legacies, different units are typically used for different parts of the spectrum. For example, radio waves are usually specified as frequencies (typically in units of MHz), while the visible region is usually specified in wavelengths (typically in units of nm or angstroms).



Portions of the electromagnetic spectrum are shown in order of decreasing frequency and increasing wavelength. (credit “Cosmic ray”: modification of work by NASA; credit “PET scan”: modification of work by the National Institute of Health; credit “X-ray”: modification of work by Dr. Jochen Lengerke; credit “Dental curing”: modification of work by the Department of the Navy; credit “Night vision”: modification of work by the Department of the Army; credit “Remote”: modification of work by Emilian Robert Vicol; credit “Cell phone”: modification of work by Brett Jordan; credit “Microwave oven”: modification of work by Billy Mabray; credit “Ultrasound”: modification of work by Jane Whitney; credit “AM radio”: modification of work by Dave Clausen)

Example:

Determining the Frequency and Wavelength of Radiation

A sodium streetlight gives off yellow light that has a wavelength of 589 nm ($1 \text{ nm} = 1 \times 10^{-9} \text{ m}$). What is the frequency of this light?

Solution

We can rearrange the equation $c = \lambda\nu$ to solve for the frequency:

Equation:

$$\nu = \frac{c}{\lambda}$$

Since c is expressed in meters per second, we must also convert 589 nm to meters.

Equation:

$$\nu = \left(\frac{2.998 \times 10^8 \cancel{\text{m}} \text{s}^{-1}}{589 \cancel{\text{nm}}} \right) \left(\frac{1 \times 10^9 \cancel{\text{nm}}}{1 \cancel{\text{m}}} \right) = 5.09 \times 10^{14} \text{s}^{-1}$$

Check Your Learning

One of the frequencies used to transmit and receive cellular telephone signals in the United States is 850 MHz. What is the wavelength in meters of these radio waves?

Note:

Answer:

$$0.353 \text{ m} = 35.3 \text{ cm}$$

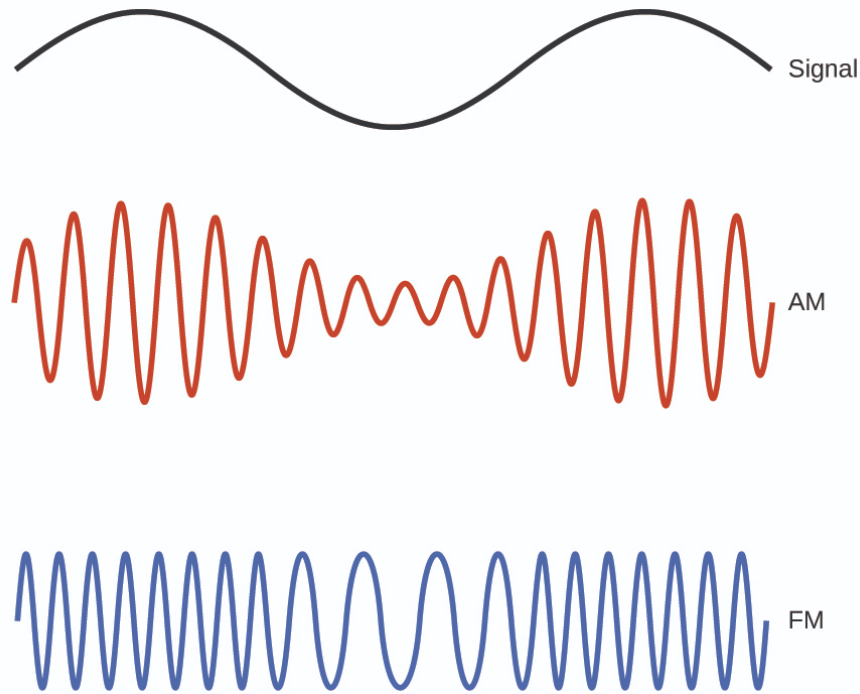
Note:

Wireless Communication



Radio and cell towers are typically used to transmit long-wavelength electromagnetic radiation. Increasingly, cell towers are designed to blend in with the landscape, as with the Tucson, Arizona, cell tower (right) disguised as a palm tree. (credit left: modification of work by Sir Mildred Pierce; credit middle: modification of work by M.O. Stevens)

Many valuable technologies operate in the radio (3 kHz-300 GHz) frequency region of the electromagnetic spectrum. At the low frequency (low energy, long wavelength) end of this region are AM (amplitude modulation) radio signals (540-2830 kHz) that can travel long distances. FM (frequency modulation) radio signals are used at higher frequencies (87.5-108.0 MHz). In AM radio, the information is transmitted by varying the amplitude of the wave ([link](#)). In FM radio, by contrast, the amplitude is constant and the instantaneous frequency varies.

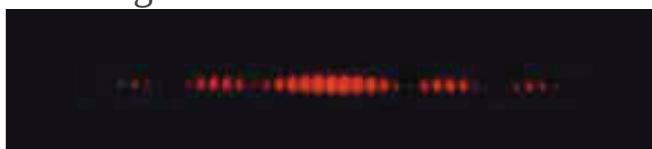


This schematic depicts how amplitude modulation (AM) and frequency modulation (FM) can be used to transmit a radio wave.

Other technologies also operate in the radio-wave portion of the electromagnetic spectrum. For example, 4G cellular telephone signals are approximately 880 MHz, while Global Positioning System (GPS) signals operate at 1.228 and 1.575 GHz, local area wireless technology (Wi-Fi) networks operate at 2.4 to 5 GHz, and highway toll sensors operate at 5.8 GHz. The frequencies associated with these applications are convenient because such waves tend not to be absorbed much by common building materials.

One particularly characteristic phenomenon of waves results when two or more waves come into contact: They interfere with each other. [\[link\]](#) shows the **interference patterns** that arise when light passes through narrow slits closely spaced about a wavelength apart. The fringe patterns produced depend on the wavelength, with the fringes being more closely spaced for shorter wavelength light passing through a given set of slits. When the light

passes through the two slits, each slit effectively acts as a new source, resulting in two closely spaced waves coming into contact at the detector (the camera in this case). The dark regions in [\[link\]](#) correspond to regions where the peaks for the wave from one slit happen to coincide with the troughs for the wave from the other slit (destructive interference), while the brightest regions correspond to the regions where the peaks for the two waves (or their two troughs) happen to coincide (constructive interference). Likewise, when two stones are tossed close together into a pond, interference patterns are visible in the interactions between the waves produced by the stones. Such interference patterns cannot be explained by particles moving according to the laws of classical mechanics.



Interference fringe patterns are shown for light passing through two closely spaced, narrow slits. The spacing of the fringes depends on the wavelength, with the fringes being more closely spaced for the shorter-wavelength blue light.
(credit: PASCO)

Note:

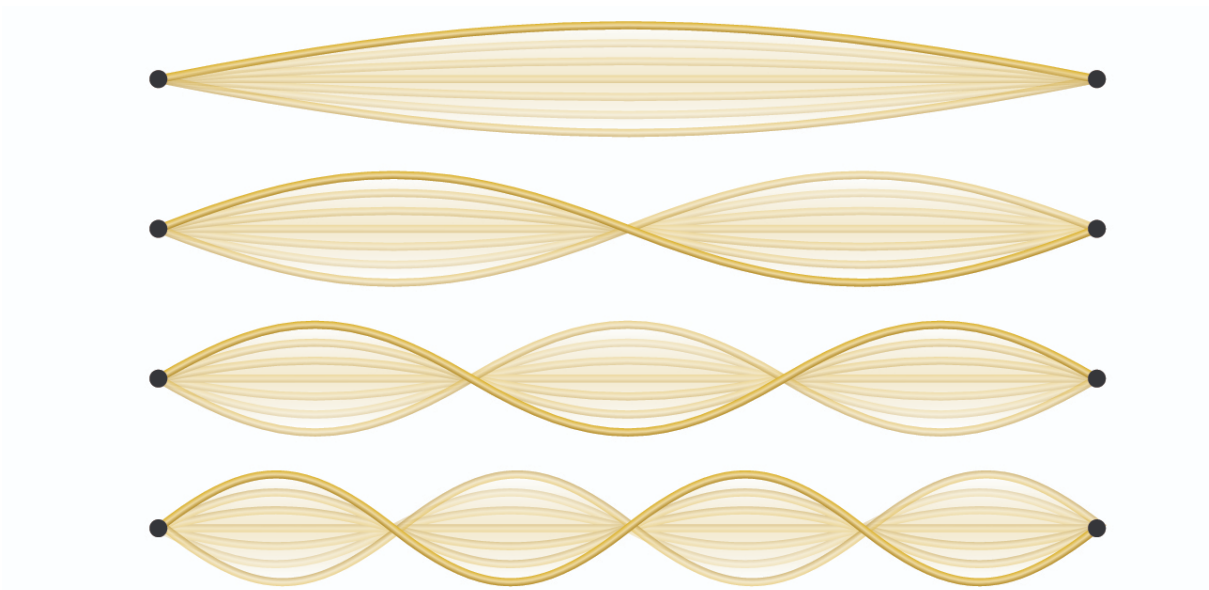
Dorothy Hodgkin

Because the wavelengths of X-rays (10-10,000 picometers [pm]) are comparable to the size of atoms, X-rays can be used to determine the structure of molecules. When a beam of X-rays is passed through molecules packed together in a crystal, the X-rays collide with the electrons and scatter. Constructive and destructive interference of these scattered X-rays creates a specific diffraction pattern. Calculating backward from this pattern, the positions of each of the atoms in the

molecule can be determined very precisely. One of the pioneers who helped create this technology was Dorothy Crowfoot Hodgkin. She was born in Cairo, Egypt, in 1910, where her British parents were studying archeology. Even as a young girl, she was fascinated with minerals and crystals. When she was a student at Oxford University, she began researching how X-ray crystallography could be used to determine the structure of biomolecules. She invented new techniques that allowed her and her students to determine the structures of vitamin B₁₂, penicillin, and many other important molecules. Diabetes, a disease that affects 382 million people worldwide, involves the hormone insulin. Hodgkin began studying the structure of insulin in 1934, but it required several decades of advances in the field before she finally reported the structure in 1969. Understanding the structure has led to better understanding of the disease and treatment options.

Not all waves are travelling waves. **Standing waves** (also known as **stationary waves**) remain constrained within some region of space. As we shall see, standing waves play an important role in our understanding of the electronic structure of atoms and molecules. The simplest example of a standing wave is a one-dimensional wave associated with a vibrating string that is held fixed at its two end points. [\[link\]](#) shows the four lowest-energy standing waves (the fundamental wave and the lowest three harmonics) for a vibrating string at a particular amplitude. Although the string's motion lies mostly within a plane, the wave itself is considered to be one dimensional, since it lies along the length of the string. The motion of string segments in a direction perpendicular to the string length generates the waves and so the amplitude of the waves is visible as the maximum displacement of the curves seen in [\[link\]](#). The key observation from the figure is *that only those waves having an integer number, n , of half-wavelengths between the end points can form*. A system with fixed end points such as this restricts the number and type of the possible waveforms. This is an example of **quantization**, in which only discrete values from a more general set of continuous values of some property are observed. Another important observation is that the harmonic waves (those waves displaying more than one-half wavelength) all have one or more points between the two end

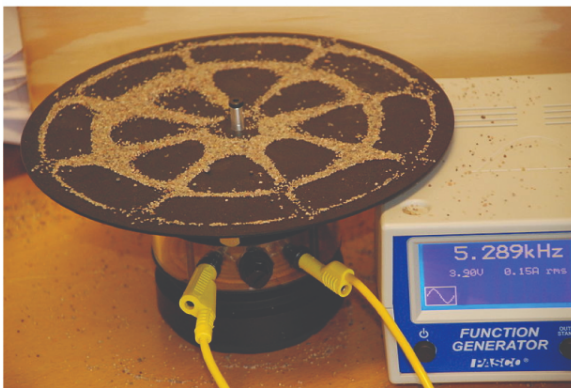
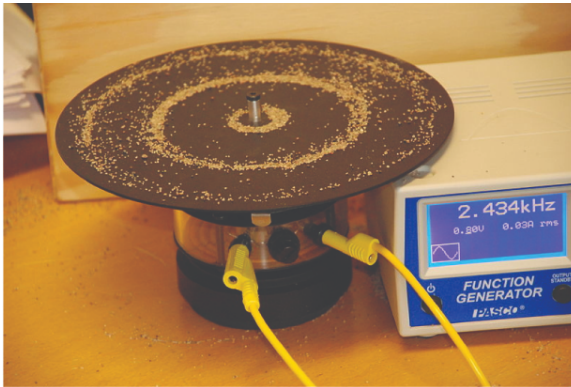
points that are not in motion. These special points are **nodes**. The energies of the standing waves with a given amplitude in a vibrating string increase with the number of half-wavelengths n . Since the number of nodes is $n - 1$, the energy can also be said to depend on the number of nodes, generally increasing as the number of nodes increases.



A vibrating string shows some one-dimensional standing waves. Since the two end points of the string are held fixed, only waves having an integer number of half-wavelengths can form. The points on the string between the end points that are not moving are called the nodes.

An example of two-dimensional standing waves is shown in [\[link\]](#), which shows the vibrational patterns on a flat surface. Although the vibrational amplitudes cannot be seen like they could in the vibrating string, the nodes have been made visible by sprinkling the drum surface with a powder that collects on the areas of the surface that have minimal displacement. For one-dimensional standing waves, the nodes were points on the line, but for two-dimensional standing waves, the nodes are lines on the surface (for

three-dimensional standing waves, the nodes are two-dimensional surfaces within the three-dimensional volume).



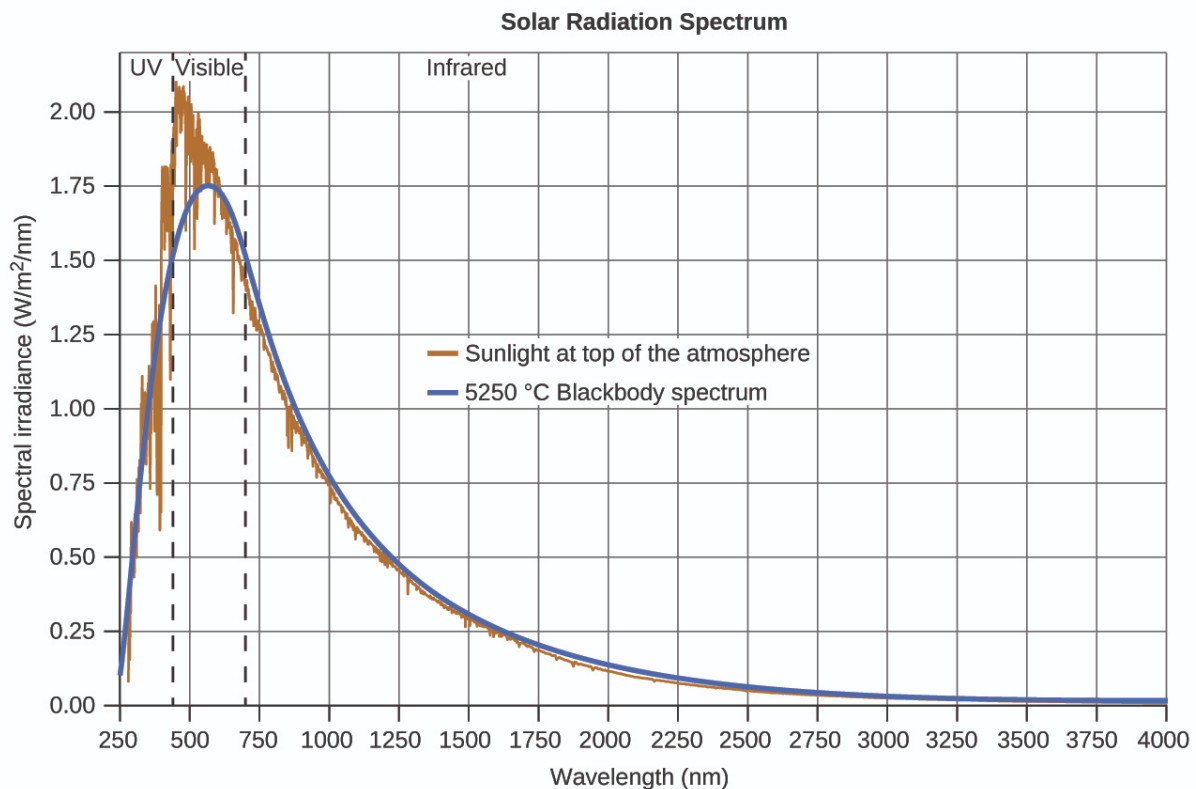
Two-dimensional standing waves can be visualized on a vibrating surface. The surface has been sprinkled with a powder that collects near the nodal lines. There are two types of nodes visible: radial nodes (circles) and angular nodes (radii).

Note:

You can watch the formation of various radial nodes [here](#) as singer Imogen Heap projects her voice across a kettle drum.

Blackbody Radiation and the Ultraviolet Catastrophe

The last few decades of the nineteenth century witnessed intense research activity in commercializing newly discovered electric lighting. This required obtaining a better understanding of the distributions of light emitted from various sources being considered. Artificial lighting is usually designed to mimic natural sunlight within the limitations of the underlying technology. Such lighting consists of a range of broadly distributed frequencies that form a **continuous spectrum**. [\[link\]](#) shows the wavelength distribution for sunlight. The most intense radiation is in the visible region, with the intensity dropping off rapidly for shorter wavelength ultraviolet (UV) light, and more slowly for longer wavelength infrared (IR) light.



The spectral distribution (light intensity vs. wavelength) of sunlight reaches the Earth's atmosphere as UV light, visible light, and IR light. The unabsorbed sunlight at the top of the atmosphere has a distribution that approximately matches the theoretical distribution of a blackbody

at 5250 °C, represented by the blue curve. (credit: modification of work by American Society for Testing and Materials (ASTM) Terrestrial Reference Spectra for Photovoltaic Performance Evaluation)

In [\[link\]](#), the solar distribution is compared to a representative distribution, called a blackbody spectrum, that corresponds to a temperature of 5250 °C. The blackbody spectrum matches the solar spectrum quite well. A **blackbody** is a convenient, ideal emitter that approximates the behavior of many materials when heated. It is “ideal” in the same sense that an ideal gas is a convenient, simple representation of real gases that works well, provided that the pressure is not too high nor the temperature too low. A good approximation of a blackbody that can be used to observe blackbody radiation is a metal oven that can be heated to very high temperatures. The oven has a small hole allowing for the light being emitted within the oven to be observed with a spectrometer so that the wavelengths and their intensities can be measured. [\[link\]](#) shows the resulting curves for some representative temperatures. Each distribution depends only on a single parameter: the temperature. The maxima in the blackbody curves, λ_{max} , shift to shorter wavelengths as the temperature increases, reflecting the observation that metals being heated to high temperatures begin to glow a darker red that becomes brighter as the temperature increases, eventually becoming white hot at very high temperatures as the intensities of all of the visible wavelengths become appreciable. This common observation was at the heart of the first paradox that showed the fundamental limitations of classical physics that we will examine.

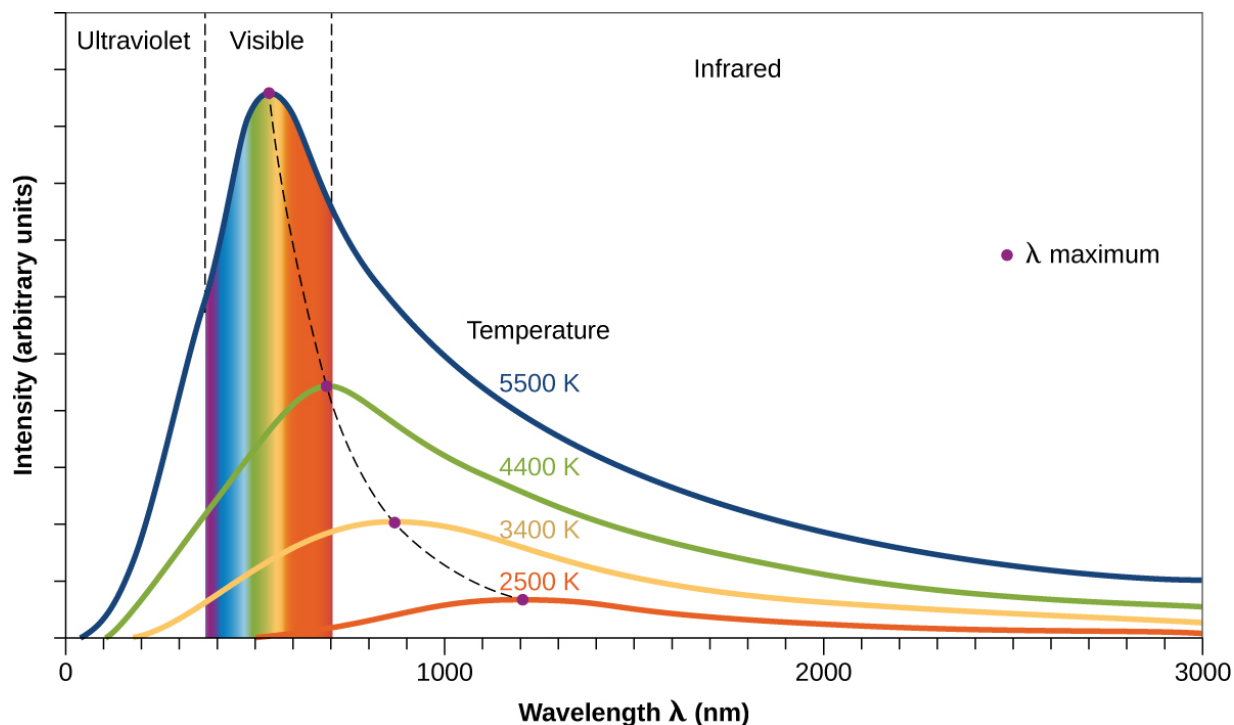
Physicists derived mathematical expressions for the blackbody curves using well-accepted concepts from the theories of classical mechanics and classical electromagnetism. The theoretical expressions as functions of temperature fit the observed experimental blackbody curves well at longer wavelengths, but showed significant discrepancies at shorter wavelengths. Not only did the theoretical curves not show a peak, they absurdly showed the intensity becoming infinitely large as the wavelength became smaller, which would imply that everyday objects at room temperature should be

emitting large amounts of UV light. This became known as the “ultraviolet catastrophe” because no one could find any problems with the theoretical treatment that could lead to such unrealistic short-wavelength behavior. Finally, around 1900, Max Planck derived a theoretical expression for blackbody radiation that fit the experimental observations exactly (within experimental error). Planck developed his theoretical treatment by extending the earlier work that had been based on the premise that the atoms composing the oven vibrated at increasing frequencies (or decreasing wavelengths) as the temperature increased, with these vibrations being the source of the emitted electromagnetic radiation. But where the earlier treatments had allowed the vibrating atoms to have any energy values obtained from a continuous set of energies (perfectly reasonable, according to classical physics), Planck found that by restricting the vibrational energies to discrete values for each frequency, he could derive an expression for blackbody radiation that correctly had the intensity dropping rapidly for the short wavelengths in the UV region.

Equation:

$$E = nh\nu, n = 1, 2, 3, \dots$$

The quantity h is a constant now known as Planck's constant, in his honor. Although Planck was pleased he had resolved the blackbody radiation paradox, he was disturbed that to do so, he needed to assume the vibrating atoms required quantized energies, which he was unable to explain. The value of Planck's constant is very small, 6.626×10^{-34} joule seconds (J s), which helps explain why energy quantization had not been observed previously in macroscopic phenomena.



Blackbody spectral distribution curves are shown for some representative temperatures.

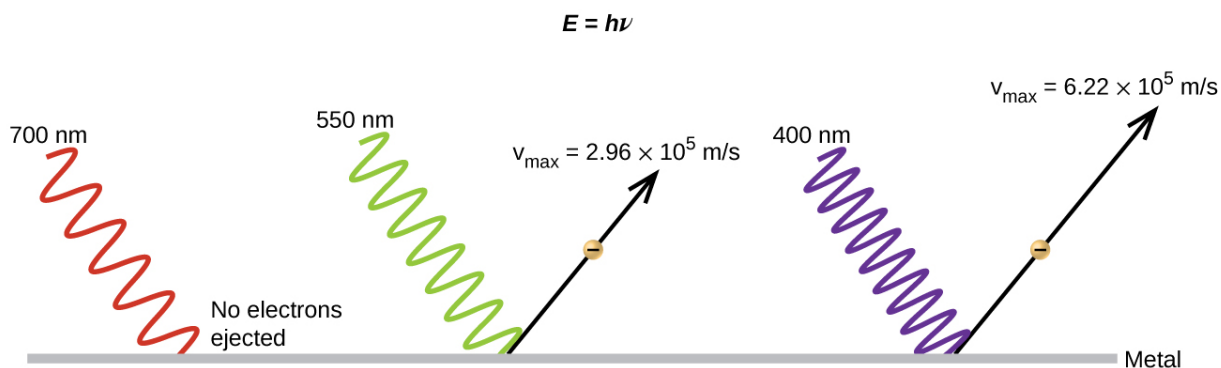
The Photoelectric Effect

The next paradox in the classical theory to be resolved concerned the photoelectric effect ([\[link\]](#)). It had been observed that electrons could be ejected from the clean surface of a metal when light having a frequency greater than some threshold frequency was shone on it. Surprisingly, the kinetic energy of the ejected electrons did not depend on the brightness of the light, but increased with increasing frequency of the light. Since the electrons in the metal had a certain amount of binding energy keeping them there, the incident light needed to have more energy to free the electrons. According to classical wave theory, a wave's energy depends on its intensity (which depends on its amplitude), not its frequency. One part of these observations was that the number of electrons ejected within in a given time period was seen to increase as the brightness increased. In 1905, Albert Einstein was able to resolve the paradox by incorporating Planck's

quantization findings into the discredited particle view of light (Einstein actually won his Nobel prize for this work, and not for his theories of relativity for which he is most famous).

Einstein argued that the quantized energies that Planck had postulated in his treatment of blackbody radiation could be applied to the light in the photoelectric effect so that the light striking the metal surface should not be viewed as a wave, but instead as a stream of particles (later called **photons**) whose energy depended on their frequency, according to Planck's formula, $E = h\nu$ (or, in terms of wavelength using $c = \nu\lambda$, $E = \frac{hc}{\lambda}$). Electrons were ejected when hit by photons having sufficient energy (a frequency greater than the threshold). The greater the frequency, the greater the kinetic energy imparted to the escaping electrons by the collisions. Einstein also argued that the light intensity did not depend on the amplitude of the incoming wave, but instead corresponded to the number of photons striking the surface within a given time period. This explains why the number of ejected electrons increased with increasing brightness, since the greater the number of incoming photons, the greater the likelihood that they would collide with some of the electrons.

With Einstein's findings, the nature of light took on a new air of mystery. Although many light phenomena could be explained either in terms of waves or particles, certain phenomena, such as the interference patterns obtained when light passed through a double slit, were completely contrary to a particle view of light, while other phenomena, such as the photoelectric effect, were completely contrary to a wave view of light. Somehow, at a deep fundamental level still not fully understood, light is both wavelike and particle-like. This is known as **wave-particle duality**.



Photons with low frequencies do not have enough energy to cause electrons to be ejected via the photoelectric effect. For any frequency of light above the threshold frequency, the kinetic energy of an ejected electron will increase linearly with the energy of the incoming photon.

Example:

Calculating the Energy of Radiation

When we see light from a neon sign, we are observing radiation from excited neon atoms. If this radiation has a wavelength of 640 nm, what is the energy of the photon being emitted?

Solution

We use the part of Planck's equation that includes the wavelength, λ , and convert units of nanometers to meters so that the units of λ and c are the same.

Equation:

$$E = \frac{hc}{\lambda}$$

Equation:

$$E = \frac{(6.626 \times 10^{-34} \text{ J}\cdot\text{s})(2.998 \times 10^8 \text{ m}\cdot\text{s}^{-1})}{(640\text{-nm})\left(\frac{1 \text{ m}}{10^9 \text{ nm}}\right)}$$

$$E = 3.10 \times 10^{-19} \text{ J}$$

Check Your Learning

The microwaves in an oven are of a specific frequency that will heat the water molecules contained in food. (This is why most plastics and glass do not become hot in a microwave oven—they do not contain water molecules.) This frequency is about 3×10^9 Hz. What is the energy of one photon in these microwaves?

Note:

Answer:

$$2 \times 10^{-24} \text{ J}$$

Note:

Use this [simulation program](#) to experiment with the photoelectric effect to see how intensity, frequency, type of metal, and other factors influence the ejected photons.

Example:

Photoelectric Effect

Identify which of the following statements are false and, where necessary, change the italicized word or phrase to make them true, consistent with Einstein's explanation of the photoelectric effect.

(a) Increasing the brightness of incoming light *increases* the kinetic energy of the ejected electrons.

- (b) Increasing the *wavelength* of incoming light increases the kinetic energy of the ejected electrons.
- (c) Increasing the brightness of incoming light *increases* the number of ejected electrons.
- (d) Increasing the *frequency* of incoming light can increase the number of ejected electrons.

Solution

- (a) False. Increasing the brightness of incoming light *has no effect* on the kinetic energy of the ejected electrons. Only energy, not the number or amplitude, of the photons influences the kinetic energy of the electrons.
- (b) False. Increasing the *frequency* of incoming light increases the kinetic energy of the ejected electrons. Frequency is proportional to energy and inversely proportional to wavelength. Frequencies above the threshold value transfer the excess energy into the kinetic energy of the electrons.
- (c) True. Because the number of collisions with photons increases with brighter light, the number of ejected electrons increases.
- (d) True with regard to the threshold energy binding the electrons to the metal. Below this threshold, electrons are not emitted and above it they are. Once over the threshold value, further increasing the frequency does not increase the number of ejected electrons

Check Your Learning

Calculate the threshold energy in kJ/mol of electrons in aluminum, given that the lowest frequency photon for which the photoelectric effect is observed is 9.87×10^{14} Hz.

Note:

Answer:

394 kJ/mol

Line Spectra

Another paradox within the classical electromagnetic theory that scientists in the late nineteenth century struggled with concerned the light emitted from atoms and molecules. When solids, liquids, or condensed gases are heated sufficiently, they radiate some of the excess energy as light. Photons produced in this manner have a range of energies, and thereby produce a continuous spectrum in which an unbroken series of wavelengths is present. Most of the light generated from stars (including our sun) is produced in this fashion. You can see all the visible wavelengths of light present in sunlight by using a prism to separate them. As can be seen in [\[link\]](#), sunlight also contains UV light (shorter wavelengths) and IR light (longer wavelengths) that can be detected using instruments but that are invisible to the human eye. Incandescent (glowing) solids such as tungsten filaments in incandescent lights also give off light that contains all wavelengths of visible light. These continuous spectra can often be approximated by blackbody radiation curves at some appropriate temperature, such as those shown in [\[link\]](#).

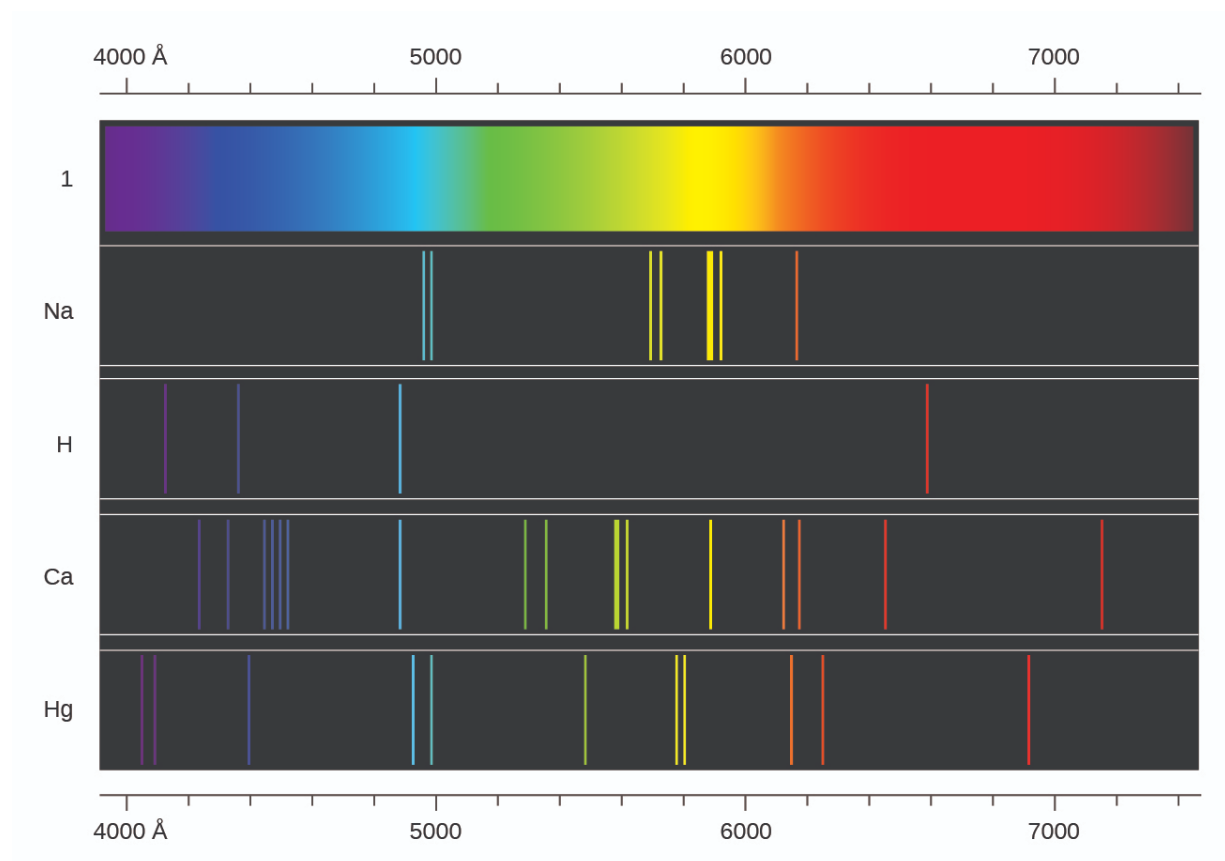
In contrast to continuous spectra, light can also occur as discrete or **line spectra** having very narrow line widths interspersed throughout the spectral regions such as those shown in [\[link\]](#). Exciting a gas at low partial pressure using an electrical current, or heating it, will produce line spectra. Fluorescent light bulbs and neon signs operate in this way ([\[link\]](#)). Each element displays its own characteristic set of lines, as do molecules, although their spectra are generally much more complicated.



Neon signs operate by exciting a gas at

low partial pressure using an electrical current. This sign shows the elaborate artistic effects that can be achieved.
(credit: Dave Shaver)

Each emission line consists of a single wavelength of light, which implies that the light emitted by a gas consists of a set of discrete energies. For example, when an electric discharge passes through a tube containing hydrogen gas at low pressure, the H_2 molecules are broken apart into separate H atoms and we see a blue-pink color. Passing the light through a prism produces a line spectrum, indicating that this light is composed of photons of four visible wavelengths, as shown in [\[link\]](#).



Compare the two types of emission spectra: continuous spectrum of

white light (top) and the line spectra of the light from excited sodium, hydrogen, calcium, and mercury atoms.

The origin of discrete spectra in atoms and molecules was extremely puzzling to scientists in the late nineteenth century, since according to classical electromagnetic theory, only continuous spectra should be observed. Even more puzzling, in 1885, Johann Balmer was able to derive an empirical equation that related the four visible wavelengths of light emitted by hydrogen atoms to whole integers. That equation is the following one, in which k is a constant:

Equation:

$$\frac{1}{\lambda} = k \left(\frac{1}{4} - \frac{1}{n^2} \right), \quad n = 3, 4, 5, 6$$

Other discrete lines for the hydrogen atom were found in the UV and IR regions. Johannes Rydberg generalized Balmer's work and developed an empirical formula that predicted all of hydrogen's emission lines, not just those restricted to the visible range, where, n_1 and n_2 are integers, $n_1 < n_2$, and R_∞ is the Rydberg constant ($1.097 \times 10^7 \text{ m}^{-1}$).

Equation:

$$\frac{1}{\lambda} = R_\infty \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right)$$

Even in the late nineteenth century, spectroscopy was a very precise science, and so the wavelengths of hydrogen were measured to very high accuracy, which implied that the Rydberg constant could be determined very precisely as well. That such a simple formula as the Rydberg formula could account for such precise measurements seemed astounding at the time, but it was the eventual explanation for emission spectra by Neils Bohr in 1913 that ultimately convinced scientists to abandon classical physics and spurred the development of modern quantum mechanics.

Key Concepts and Summary

Light and other forms of electromagnetic radiation move through a vacuum with a constant speed, c , of $2.998 \times 10^8 \text{ m s}^{-1}$. This radiation shows wavelike behavior, which can be characterized by a frequency, ν , and a wavelength, λ , such that $c = \lambda\nu$. Light is an example of a travelling wave. Other important wave phenomena include standing waves, periodic oscillations, and vibrations. Standing waves exhibit quantization, since their wavelengths are limited to discrete integer multiples of some characteristic lengths. Electromagnetic radiation that passes through two closely spaced narrow slits having dimensions roughly similar to the wavelength will show an interference pattern that is a result of constructive and destructive interference of the waves. Electromagnetic radiation also demonstrates properties of particles called photons. The energy of a photon is related to the frequency (or alternatively, the wavelength) of the radiation as $E = h\nu$ (or $E = \frac{hc}{\lambda}$), where h is Planck's constant. That light demonstrates both wavelike and particle-like behavior is known as wave-particle duality. All forms of electromagnetic radiation share these properties, although various forms including X-rays, visible light, microwaves, and radio waves interact differently with matter and have very different practical applications. Electromagnetic radiation can be generated by exciting matter to higher energies, such as by heating it. The emitted light can be either continuous (incandescent sources like the sun) or discrete (from specific types of excited atoms). Continuous spectra often have distributions that can be approximated as blackbody radiation at some appropriate temperature. The line spectrum of hydrogen can be obtained by passing the light from an electrified tube of hydrogen gas through a prism. This line spectrum was simple enough that an empirical formula called the Rydberg formula could be derived from the spectrum. Three historically important paradoxes from the late 19th and early 20th centuries that could not be explained within the existing framework of classical mechanics and classical electromagnetism were the blackbody problem, the photoelectric effect, and the discrete spectra of atoms. The resolution of these paradoxes ultimately led to quantum theories that superseded the classical theories.

Key Equations

- $c = \lambda\nu$
- $E = h\nu = \frac{hc}{\lambda}$, where $h = 6.626 \times 10^{-34} \text{ J s}$
- $\frac{1}{\lambda} = R_{\infty} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right)$

Chemistry End of Chapter Exercises

Exercise:

Problem:

The light produced by a red neon sign is due to the emission of light by excited neon atoms. Qualitatively describe the spectrum produced by passing light from a neon lamp through a prism.

Solution:

The spectrum consists of colored lines, at least one of which (probably the brightest) is red.

Exercise:

Problem:

An FM radio station found at 103.1 on the FM dial broadcasts at a frequency of $1.031 \times 10^8 \text{ s}^{-1}$ (103.1 MHz). What is the wavelength of these radio waves in meters?

Exercise:

Problem:

FM-95, an FM radio station, broadcasts at a frequency of $9.51 \times 10^7 \text{ s}^{-1}$ (95.1 MHz). What is the wavelength of these radio waves in meters?

Solution:

3.15 m

Exercise:

Problem:

A bright violet line occurs at 435.8 nm in the emission spectrum of mercury vapor. What amount of energy, in joules, must be released by an electron in a mercury atom to produce a photon of this light?

Exercise:**Problem:**

Light with a wavelength of 614.5 nm looks orange. What is the energy, in joules, per photon of this orange light? What is the energy in eV ($1 \text{ eV} = 1.602 \times 10^{-19} \text{ J}$)?

Solution:

$$3.233 \times 10^{-19} \text{ J}; 2.018 \text{ eV}$$

Exercise:**Problem:**

Heated lithium atoms emit photons of light with an energy of $2.961 \times 10^{-19} \text{ J}$. Calculate the frequency and wavelength of one of these photons. What is the total energy in 1 mole of these photons? What is the color of the emitted light?

Exercise:**Problem:**

A photon of light produced by a surgical laser has an energy of $3.027 \times 10^{-19} \text{ J}$. Calculate the frequency and wavelength of the photon. What is the total energy in 1 mole of photons? What is the color of the emitted light?

Solution:

$$\nu = 4.568 \times 10^{14} \text{ s}^{-1}; \lambda = 656.3 \text{ nm}; \text{Energy mol}^{-1} = 1.823 \times 10^5 \text{ J mol}^{-1}; \text{red}$$

Exercise:**Problem:**

When rubidium ions are heated to a high temperature, two lines are observed in its line spectrum at wavelengths (a) 7.9×10^{-7} m and (b) 4.2×10^{-7} m. What are the frequencies of the two lines? What color do we see when we heat a rubidium compound?

Exercise:**Problem:**

The emission spectrum of cesium contains two lines whose frequencies are (a) 3.45×10^{14} Hz and (b) 6.53×10^{14} Hz. What are the wavelengths and energies per photon of the two lines? What color are the lines?

Solution:

(a) $\lambda = 8.69 \times 10^{-7}$ m; $E = 2.29 \times 10^{-19}$ J; (b) $\lambda = 4.59 \times 10^{-7}$ m; $E = 4.33 \times 10^{-19}$ J; The color of (a) is red; (b) is blue.

Exercise:**Problem:**

Photons of infrared radiation are responsible for much of the warmth we feel when holding our hands before a fire. These photons will also warm other objects. How many infrared photons with a wavelength of 1.5×10^{-6} m must be absorbed by the water to warm a cup of water (175 g) from 25.0 °C to 40 °C?

Exercise:**Problem:**

One of the radiographic devices used in a dentist's office emits an X-ray of wavelength 2.090×10^{-11} m. What is the energy, in joules, and frequency of this X-ray?

Solution:

$$E = 9.502 \times 10^{-15} \text{ J}; \nu = 1.434 \times 10^{19} \text{ s}^{-1}$$

Exercise:**Problem:**

The eyes of certain reptiles pass a single visual signal to the brain when the visual receptors are struck by photons of a wavelength of 850 nm. If a total energy of $3.15 \times 10^{-14} \text{ J}$ is required to trip the signal, what is the minimum number of photons that must strike the receptor?

Exercise:**Problem:**

RGB color television and computer displays use cathode ray tubes that produce colors by mixing red, green, and blue light. If we look at the screen with a magnifying glass, we can see individual dots turn on and off as the colors change. Using a spectrum of visible light, determine the approximate wavelength of each of these colors. What is the frequency and energy of a photon of each of these colors?

Solution:

Red: 660 nm; $4.54 \times 10^{14} \text{ Hz}$; $3.01 \times 10^{-19} \text{ J}$. Green: 520 nm; $5.77 \times 10^{14} \text{ Hz}$; $3.82 \times 10^{-19} \text{ J}$. Blue: 440 nm; $6.81 \times 10^{14} \text{ Hz}$; $4.51 \times 10^{-19} \text{ J}$. Somewhat different numbers are also possible.

Exercise:

Problem: Answer the following questions about a Blu-ray laser:

(a) The laser on a Blu-ray player has a wavelength of 405 nm. In what region of the electromagnetic spectrum is this radiation? What is its frequency?

(b) A Blu-ray laser has a power of 5 milliwatts ($1 \text{ watt} = 1 \text{ J s}^{-1}$). How many photons of light are produced by the laser in 1 hour?

(c) The ideal resolution of a player using a laser (such as a Blu-ray player), which determines how close together data can be stored on a compact disk, is determined using the following formula: Resolution = $0.60(\lambda/\text{NA})$, where λ is the wavelength of the laser and NA is the numerical aperture. Numerical aperture is a measure of the size of the spot of light on the disk; the larger the NA, the smaller the spot. In a typical Blu-ray system, $\text{NA} = 0.95$. If the 405-nm laser is used in a Blu-ray player, what is the closest that information can be stored on a Blu-ray disk?

(d) The data density of a Blu-ray disk using a 405-nm laser is 1.5×10^7 bits mm^{-2} . Disks have an outside diameter of 120 mm and a hole of 15-mm diameter. How many data bits can be contained on the disk? If a Blu-ray disk can hold 9,400,000 pages of text, how many data bits are needed for a typed page? (Hint: Determine the area of the disk that is available to hold data. The area inside a circle is given by $A = \pi r^2$, where the radius r is one-half of the diameter.)

Exercise:

Problem:

What is the threshold frequency for sodium metal if a photon with frequency $6.66 \times 10^{14} \text{ s}^{-1}$ ejects an electron with $7.74 \times 10^{-20} \text{ J}$ kinetic energy? Will the photoelectric effect be observed if sodium is exposed to orange light?

Solution:

$5.49 \times 10^{14} \text{ s}^{-1}$; no

Glossary

amplitude

extent of the displacement caused by a wave

blackbody

idealized perfect absorber of all incident electromagnetic radiation;
such bodies emit electromagnetic radiation in characteristic continuous
spectra called blackbody radiation

continuous spectrum

electromagnetic radiation given off in an unbroken series of
wavelengths (e.g., white light from the sun)

electromagnetic radiation

energy transmitted by waves that have an electric-field component and
a magnetic-field component

electromagnetic spectrum

range of energies that electromagnetic radiation can comprise,
including radio, microwaves, infrared, visible, ultraviolet, X-rays, and
gamma rays

frequency (ν)

number of wave cycles (peaks or troughs) that pass a specified point in
space per unit time

hertz (Hz)

the unit of frequency, which is the number of cycles per second, s^{-1}

intensity

property of wave-propagated energy related to the amplitude of the
wave, such as brightness of light or loudness of sound

interference pattern

pattern typically consisting of alternating bright and dark fringes; it
results from constructive and destructive interference of waves

line spectrum

electromagnetic radiation emitted at discrete wavelengths by a specific
atom (or atoms) in an excited state

node

any point of a standing wave with zero amplitude

photon

smallest possible packet of electromagnetic radiation, a particle of light

quantization

limitation of some property to specific discrete values, not continuous

standing wave

(also, stationary wave) localized wave phenomenon characterized by discrete wavelengths determined by the boundary conditions used to generate the waves; standing waves are inherently quantized

wave

oscillation of a property over time or space; can transport energy from one point to another

wavelength (λ)

distance between two consecutive peaks or troughs in a wave

wave-particle duality

observation that elementary particles can exhibit both wave-like and particle-like properties

The Bohr Model

By the end of this section, you will be able to:

- Describe the Bohr model of the hydrogen atom
- Use the Rydberg equation to calculate energies of light emitted or absorbed by hydrogen atoms

Following the work of Ernest Rutherford and his colleagues in the early twentieth century, the picture of atoms consisting of tiny dense nuclei surrounded by lighter and even tinier electrons continually moving about the nucleus was well established. This picture was called the planetary model, since it pictured the atom as a miniature “solar system” with the electrons orbiting the nucleus like planets orbiting the sun. The simplest atom is hydrogen, consisting of a single proton as the nucleus about which a single electron moves. The electrostatic force attracting the electron to the proton depends only on the distance between the two particles. This classical mechanics description of the atom is incomplete, however, since an electron moving in an elliptical orbit would be accelerating (by changing direction) and, according to classical electromagnetism, it should continuously emit electromagnetic radiation. This loss in orbital energy should result in the electron’s orbit getting continually smaller until it spirals into the nucleus, implying that atoms are inherently unstable.

In 1913, Niels Bohr attempted to resolve the atomic paradox by ignoring classical electromagnetism’s prediction that the orbiting electron in hydrogen would continuously emit light. Instead, he incorporated into the classical mechanics description of the atom Planck’s ideas of quantization and Einstein’s finding that light consists of photons whose energy is proportional to their frequency. Bohr assumed that the electron orbiting the nucleus would not normally emit any radiation (the stationary state hypothesis), but it would emit or absorb a photon if it moved to a different orbit. The energy absorbed or emitted would reflect differences in the orbital energies according to this equation:

Equation:

$$|\Delta E| = |E_f - E_i| = h\nu = \frac{hc}{\lambda}$$

In this equation, h is Planck's constant and E_i and E_f are the initial and final orbital energies, respectively. The absolute value of the energy difference is used, since frequencies and wavelengths are always positive. Instead of allowing for continuous values of energy, Bohr assumed the energies of these electron orbitals were quantized:

Equation:

$$E_n = -\frac{k}{n^2}, n = 1, 2, 3, \dots$$

In this expression, k is a constant comprising fundamental constants such as the electron mass and charge and Planck's constant. Inserting the expression for the orbit energies into the equation for ΔE gives

Equation:

$$\Delta E = k \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) = \frac{hc}{\lambda}$$

or

Equation:

$$\frac{1}{\lambda} = \frac{k}{hc} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right)$$

which is identical to the Rydberg equation in which $R_\infty = \frac{k}{hc}$. When Bohr calculated his theoretical value for the Rydberg constant, R_∞ , and compared it with the experimentally accepted value, he got excellent agreement. Since the Rydberg constant was one of the most precisely measured constants at that time, this level of agreement was astonishing and meant that **Bohr's model** was taken seriously, despite the many assumptions that Bohr needed to derive it.

The lowest few energy levels are shown in [\[link\]](#). One of the fundamental laws of physics is that matter is most stable with the lowest possible energy.

Thus, the electron in a hydrogen atom usually moves in the $n = 1$ orbit, the orbit in which it has the lowest energy. When the electron is in this lowest energy orbit, the atom is said to be in its **ground electronic state** (or simply ground state). If the atom receives energy from an outside source, it is possible for the electron to move to an orbit with a higher n value and the atom is now in an **excited electronic state** (or simply an excited state) with a higher energy. When an electron transitions from an excited state (higher energy orbit) to a less excited state, or ground state, the difference in energy is emitted as a photon. Similarly, if a photon is absorbed by an atom, the energy of the photon moves an electron from a lower energy orbit up to a more excited one. We can relate the energy of electrons in atoms to what we learned previously about energy. The law of conservation of energy says that we can neither create nor destroy energy. Thus, if a certain amount of external energy is required to excite an electron from one energy level to another, that same amount of energy will be liberated when the electron returns to its initial state ([\[link\]](#)).

Since Bohr's model involved only a single electron, it could also be applied to the single electron ions He^+ , Li^{2+} , Be^{3+} , and so forth, which differ from hydrogen only in their nuclear charges, and so one-electron atoms and ions are collectively referred to as hydrogen-like atoms. The energy expression for hydrogen-like atoms is a generalization of the hydrogen atom energy, in which Z is the nuclear charge (+1 for hydrogen, +2 for He, +3 for Li, and so on) and k has a value of $2.179 \times 10^{-18} \text{ J}$.

Equation:

$$E_n = -\frac{kZ^2}{n^2}$$

The sizes of the circular orbits for hydrogen-like atoms are given in terms of their radii by the following expression, in which α_0 is a constant called the Bohr radius, with a value of $5.292 \times 10^{-11} \text{ m}$:

Equation:

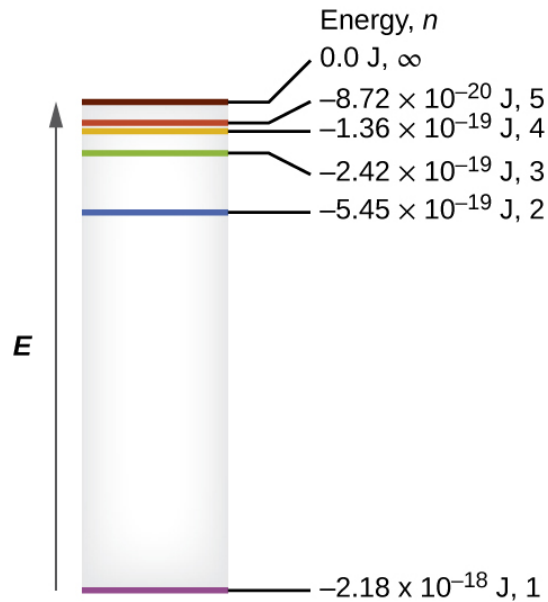
$$r = \frac{n^2}{Z} \alpha_0$$

The equation also shows us that as the electron's energy increases (as n increases), the electron is found at greater distances from the nucleus. This is implied by the inverse dependence of electrostatic attraction on distance, since, as the electron moves away from the nucleus, the electrostatic attraction between it and the nucleus decreases and it is held less tightly in the atom. Note that as n gets larger and the orbits get larger, their energies get closer to zero, and so the limits $n \rightarrow \infty$ and $r \rightarrow \infty$ imply that $E = 0$ corresponds to the ionization limit where the electron is completely removed from the nucleus. Thus, for hydrogen in the ground state $n = 1$, the ionization energy would be:

Equation:

$$\Delta E = E_{n \rightarrow \infty} - E_1 = 0 - (-13.6 \text{ eV}) = 13.6 \text{ eV}$$

With three extremely puzzling paradoxes now solved (blackbody radiation, the photoelectric effect, and the hydrogen atom), and all involving Planck's constant in a fundamental manner, it became clear to most physicists at that time that the classical theories that worked so well in the macroscopic world were fundamentally flawed and could not be extended down into the microscopic domain of atoms and molecules. Unfortunately, despite Bohr's remarkable achievement in deriving a theoretical expression for the Rydberg constant, he was unable to extend his theory to the next simplest atom, He, which only has two electrons. Bohr's model was severely flawed, since it was still based on the classical mechanics notion of precise orbits, a concept that was later found to be untenable in the microscopic domain, when a proper model of quantum mechanics was developed to supersede classical mechanics.



Quantum numbers and energy levels in a hydrogen atom. The more negative the calculated value, the lower the energy.

Example:

Calculating the Energy of an Electron in a Bohr Orbit

Early researchers were very excited when they were able to predict the energy of an electron at a particular distance from the nucleus in a hydrogen atom. If a spark promotes the electron in a hydrogen atom into an orbit with $n = 3$, what is the calculated energy, in joules, of the electron?

Solution

The energy of the electron is given by this equation:

Equation:

$$E = \frac{-kZ^2}{n^2}$$

The atomic number, Z , of hydrogen is 1; $k = 2.179 \times 10^{-18}$ J; and the electron is characterized by an n value of 3. Thus,

Equation:

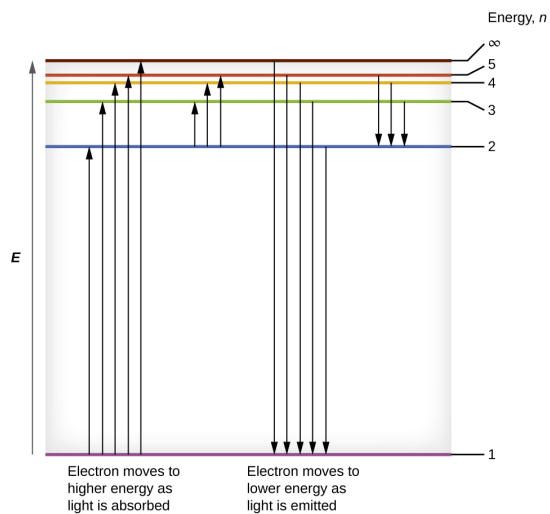
$$E = \frac{-(2.179 \times 10^{-18} \text{ J}) \times (1)^2}{(3)^2} = -2.421 \times 10^{-19} \text{ J}$$

Check Your Learning

The electron in [link](#) is promoted even further to an orbit with $n = 6$. What is its new energy?

Note:**Answer:**

$$-6.053 \times 10^{-20} \text{ J}$$



The horizontal lines show the relative energy of orbits in the Bohr model of the hydrogen atom, and the vertical arrows depict the energy of photons

absorbed (left) or emitted (right) as electrons move between these orbits.

Example:

Calculating the Energy and Wavelength of Electron Transitions in a One-electron (Bohr) System

What is the energy (in joules) and the wavelength (in meters) of the line in the spectrum of hydrogen that represents the movement of an electron from Bohr orbit with $n = 4$ to the orbit with $n = 6$? In what part of the electromagnetic spectrum do we find this radiation?

Solution

In this case, the electron starts out with $n = 4$, so $n_1 = 4$. It comes to rest in the $n = 6$ orbit, so $n_2 = 6$. The difference in energy between the two states is given by this expression:

Equation:

$$\Delta E = E_1 - E_2 = 2.179 \times 10^{-18} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right)$$

Equation:

$$\Delta E = 2.179 \times 10^{-18} \left(\frac{1}{4^2} - \frac{1}{6^2} \right) \text{ J}$$

Equation:

$$\Delta E = 2.179 \times 10^{-18} \left(\frac{1}{16} - \frac{1}{36} \right) \text{ J}$$

Equation:

$$\Delta E = 7.566 \times 10^{-20} \text{ J}$$

This energy difference is positive, indicating a photon enters the system (is absorbed) to excite the electron from the $n = 4$ orbit up to the $n = 6$ orbit. The wavelength of a photon with this energy is found by the expression $E = \frac{hc}{\lambda}$. Rearrangement gives:

Equation:

$$\lambda = \frac{hc}{E}$$

Equation:

$$\begin{aligned} &= (6.626 \times 10^{-34} \text{ J}\cdot\text{s}) \times \frac{2.998 \times 10^8 \text{ m}\cdot\text{s}^{-1}}{7.566 \times 10^{-20} \text{ J}} \\ &= 2.626 \times 10^{-6} \text{ m} \end{aligned}$$

From the illustration of the electromagnetic spectrum in Electromagnetic Energy, we can see that this wavelength is found in the infrared portion of the electromagnetic spectrum.

Check Your Learning

What is the energy in joules and the wavelength in meters of the photon produced when an electron falls from the $n = 5$ to the $n = 3$ level in a He^+ ion ($Z = 2$ for He^+)?

Note:

Answer:

$$6.198 \times 10^{-19} \text{ J}; 3.205 \times 10^{-7} \text{ m}$$

Bohr's model of the hydrogen atom provides insight into the behavior of matter at the microscopic level, but it does not account for electron–electron interactions in atoms with more than one electron. It does introduce

several important features of all models used to describe the distribution of electrons in an atom. These features include the following:

- The energies of electrons (energy levels) in an atom are quantized, described by **quantum numbers**: integer numbers having only specific allowed value and used to characterize the arrangement of electrons in an atom.
- An electron's energy increases with increasing distance from the nucleus.
- The discrete energies (lines) in the spectra of the elements result from quantized electronic energies.

Of these features, the most important is the postulate of quantized energy levels for an electron in an atom. As a consequence, the model laid the foundation for the quantum mechanical model of the atom. Bohr won a Nobel Prize in Physics for his contributions to our understanding of the structure of atoms and how that is related to line spectra emissions.

Key Concepts and Summary

Bohr incorporated Planck's and Einstein's quantization ideas into a model of the hydrogen atom that resolved the paradox of atom stability and discrete spectra. The Bohr model of the hydrogen atom explains the connection between the quantization of photons and the quantized emission from atoms. Bohr described the hydrogen atom in terms of an electron moving in a circular orbit about a nucleus. He postulated that the electron was restricted to certain orbits characterized by discrete energies. Transitions between these allowed orbits result in the absorption or emission of photons. When an electron moves from a higher-energy orbit to a more stable one, energy is emitted in the form of a photon. To move an electron from a stable orbit to a more excited one, a photon of energy must be absorbed. Using the Bohr model, we can calculate the energy of an electron and the radius of its orbit in any one-electron system.

Key Equations

- $E_n = -\frac{kZ^2}{n^2}$, $n = 1, 2, 3, \dots$
- $\Delta E = kZ^2 \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right)$
- $r = \frac{n^2}{Z} a_0$

Chemistry End of Chapter Exercises

Exercise:

Problem:

Why is the electron in a Bohr hydrogen atom bound less tightly when it has a quantum number of 3 than when it has a quantum number of 1?

Exercise:

Problem:

What does it mean to say that the energy of the electrons in an atom is quantized?

Solution:

Quantized energy means that the electrons can possess only certain discrete energy values; values between those quantized values are not permitted.

Exercise:

Problem:

Using the Bohr model, determine the energy, in joules, necessary to ionize a ground-state hydrogen atom. Show your calculations.

Exercise:

Problem:

The electron volt (eV) is a convenient unit of energy for expressing atomic-scale energies. It is the amount of energy that an electron gains when subjected to a potential of 1 volt; $1 \text{ eV} = 1.602 \times 10^{-19} \text{ J}$. Using the Bohr model, determine the energy, in electron volts, of the photon produced when an electron in a hydrogen atom moves from the orbit with $n = 5$ to the orbit with $n = 2$. Show your calculations.

Solution:

2.856 eV

Exercise:**Problem:**

Using the Bohr model, determine the lowest possible energy, in joules, for the electron in the Li^{2+} ion.

Exercise:**Problem:**

Using the Bohr model, determine the lowest possible energy for the electron in the He^+ ion.

Solution:

$-8.716 \times 10^{-18} \text{ J}$

Exercise:**Problem:**

Using the Bohr model, determine the energy of an electron with $n = 6$ in a hydrogen atom.

Exercise:

Problem:

Using the Bohr model, determine the energy of an electron with $n = 8$ in a hydrogen atom.

Solution:

$$-3.405 \times 10^{-20} \text{ J}$$

Exercise:**Problem:**

How far from the nucleus in angstroms ($1 \text{ angstrom} = 1 \times 10^{-10} \text{ m}$) is the electron in a hydrogen atom if it has an energy of $-8.72 \times 10^{-20} \text{ J}$?

Exercise:**Problem:**

What is the radius, in angstroms, of the orbital of an electron with $n = 8$ in a hydrogen atom?

Solution:

$$33.9 \text{ \AA}$$

Exercise:**Problem:**

Using the Bohr model, determine the energy in joules of the photon produced when an electron in a He^+ ion moves from the orbit with $n = 5$ to the orbit with $n = 2$.

Exercise:

Problem:

Using the Bohr model, determine the energy in joules of the photon produced when an electron in a Li^{2+} ion moves from the orbit with $n = 2$ to the orbit with $n = 1$.

Solution:

$$1.471 \times 10^{-17} \text{ J}$$

Exercise:**Problem:**

Consider a large number of hydrogen atoms with electrons randomly distributed in the $n = 1, 2, 3$, and 4 orbits.

- (a) How many different wavelengths of light are emitted by these atoms as the electrons fall into lower-energy orbitals?
- (b) Calculate the lowest and highest energies of light produced by the transitions described in part (a).
- (c) Calculate the frequencies and wavelengths of the light produced by the transitions described in part (b).

Exercise:**Problem:**

How are the Bohr model and the Rutherford model of the atom similar? How are they different?

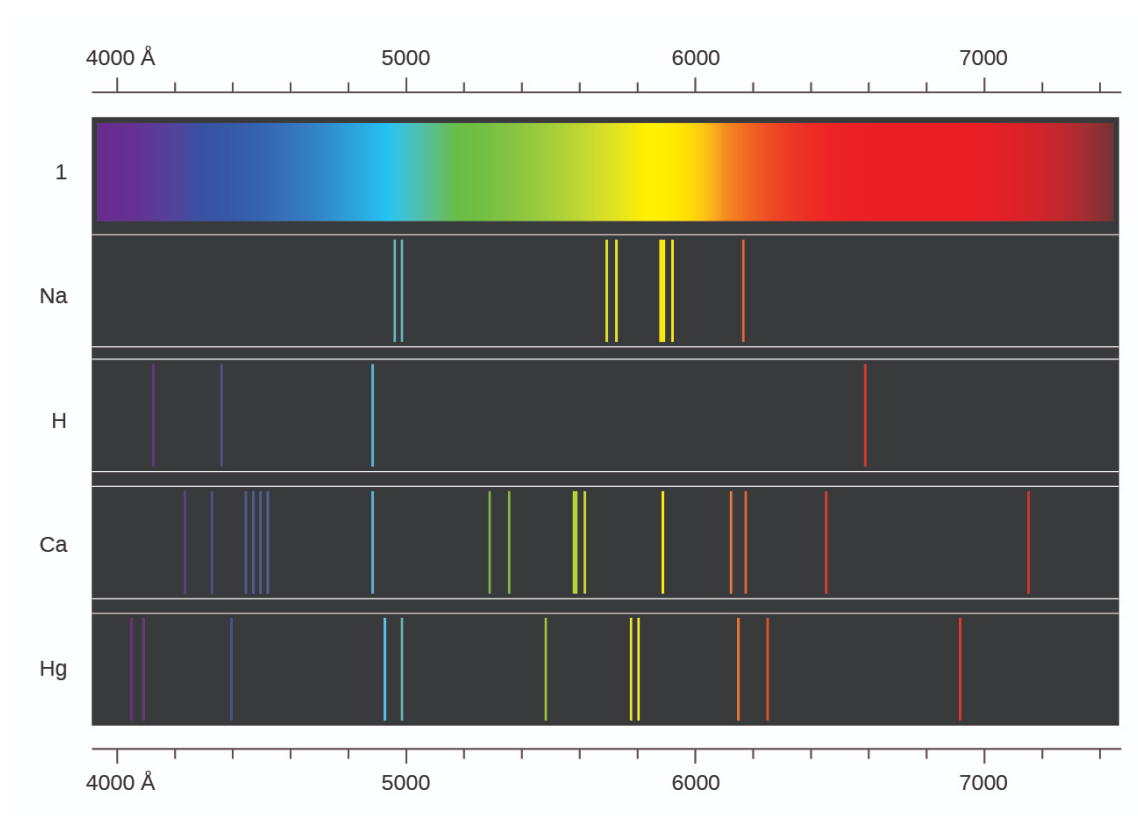
Solution:

Both involve a relatively heavy nucleus with electrons moving around it, although strictly speaking, the Bohr model works only for one-electron atoms or ions. According to classical mechanics, the Rutherford model predicts a miniature “solar system” with electrons moving about the nucleus in circular or elliptical orbits that are

confined to planes. If the requirements of classical electromagnetic theory that electrons in such orbits would emit electromagnetic radiation are ignored, such atoms would be stable, having constant energy and angular momentum, but would not emit any visible light (contrary to observation). If classical electromagnetic theory is applied, then the Rutherford atom would emit electromagnetic radiation of continually increasing frequency (contrary to the observed discrete spectra), thereby losing energy until the atom collapsed in an absurdly short time (contrary to the observed long-term stability of atoms). The Bohr model retains the classical mechanics view of circular orbits confined to planes having constant energy and angular momentum, but restricts these to quantized values dependent on a single quantum number, n . The orbiting electron in Bohr's model is assumed not to emit any electromagnetic radiation while moving about the nucleus in its stationary orbits, but the atom can emit or absorb electromagnetic radiation when the electron changes from one orbit to another. Because of the quantized orbits, such "quantum jumps" will produce discrete spectra, in agreement with observations.

Exercise:

Problem: The spectra of hydrogen and of calcium are shown here.



What causes the lines in these spectra? Why are the colors of the lines different? Suggest a reason for the observation that the spectrum of calcium is more complicated than the spectrum of hydrogen.

Glossary

Bohr's model of the hydrogen atom

structural model in which an electron moves around the nucleus only in circular orbits, each with a specific allowed radius

excited state

state having an energy greater than the ground-state energy

ground state

state in which the electrons in an atom, ion, or molecule have the lowest energy possible

quantum number

number having only specific allowed values and used to characterize the arrangement of electrons in an atom

Development of Quantum Theory

By the end of this section, you will be able to:

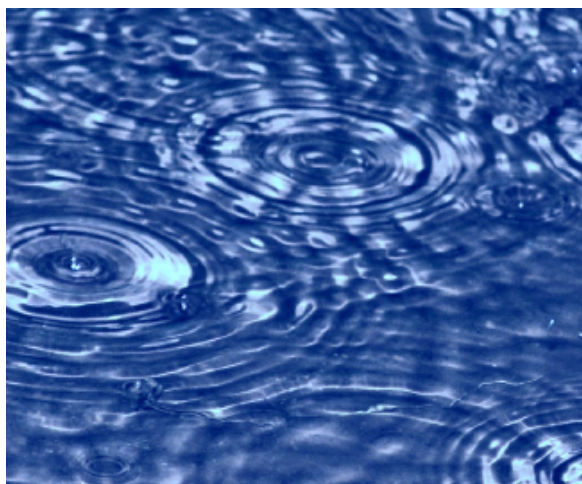
- Extend the concept of wave–particle duality that was observed in electromagnetic radiation to matter as well
- Understand the general idea of the quantum mechanical description of electrons in an atom, and that it uses the notion of three-dimensional wave functions, or orbitals, that define the distribution of probability to find an electron in a particular part of space
- List and describe traits of the four quantum numbers that form the basis for completely specifying the state of an electron in an atom

Bohr's model explained the experimental data for the hydrogen atom and was widely accepted, but it also raised many questions. Why did electrons orbit at only fixed distances defined by a single quantum number $n = 1, 2, 3$, and so on, but never in between? Why did the model work so well describing hydrogen and one-electron ions, but could not correctly predict the emission spectrum for helium or any larger atoms? To answer these questions, scientists needed to completely revise the way they thought about matter.

Behavior in the Microscopic World

We know how matter behaves in the macroscopic world—objects that are large enough to be seen by the naked eye follow the rules of classical physics. A billiard ball moving on a table will behave like a particle: It will continue in a straight line unless it collides with another ball or the table cushion, or is acted on by some other force (such as friction). The ball has a well-defined position and velocity (or a well-defined momentum, $p = mv$, defined by mass m and velocity v) at any given moment. In other words, the ball is moving in a classical trajectory. This is the typical behavior of a classical object.

When waves interact with each other, they show interference patterns that are not displayed by macroscopic particles such as the billiard ball. For example, interacting waves on the surface of water can produce interference patterns similar to those shown on [\[link\]](#). This is a case of wave behavior on the macroscopic scale, and it is clear that particles and waves are very different phenomena in the macroscopic realm.



An interference pattern on the water surface is formed by interacting waves. The waves are caused by reflection of water from the rocks. (credit: modification of work by Sukanto Debnath)

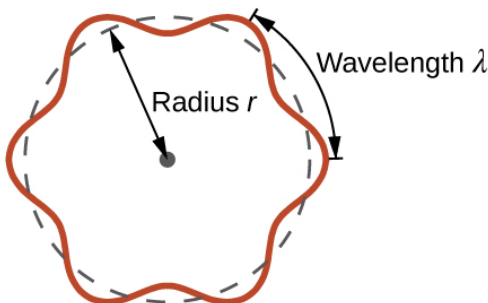
As technological improvements allowed scientists to probe the microscopic world in greater detail, it became increasingly clear by the 1920s that very small pieces of matter follow a different set of rules from those we observe for large objects. The unquestionable separation of waves and particles was no longer the case for the microscopic world.

One of the first people to pay attention to the special behavior of the microscopic world was Louis de Broglie. He asked the question: If electromagnetic radiation can have particle-like character, can electrons and other submicroscopic particles exhibit wavelike character? In his 1925 doctoral dissertation, de Broglie extended the wave-particle duality of light that Einstein used to resolve the photoelectric-effect paradox to material particles. He predicted that a particle with mass m and velocity v (that is, with linear momentum p) should also exhibit the behavior of a wave with a wavelength value λ , given by this expression in which h is the familiar Planck's constant:

Equation:

$$\lambda = \frac{h}{mv} = \frac{h}{p}$$

This is called the *de Broglie wavelength*. Unlike the other values of λ discussed in this chapter, the de Broglie wavelength is a characteristic of particles and other bodies, not electromagnetic radiation (note that this equation involves velocity [v , m/s], not frequency [ν , Hz]. Although these two symbols appear nearly identical, they mean very different things). Where Bohr had postulated the electron as being a particle orbiting the nucleus in quantized orbits, de Broglie argued that Bohr's assumption of quantization can be explained if the electron is considered not as a particle, but rather as a circular standing wave such that only an integer number of wavelengths could fit exactly within the orbit ([\[link\]](#)).



If an electron is viewed as a wave circling around the nucleus, an integer number of wavelengths must fit into the orbit for this standing wave behavior to be possible.

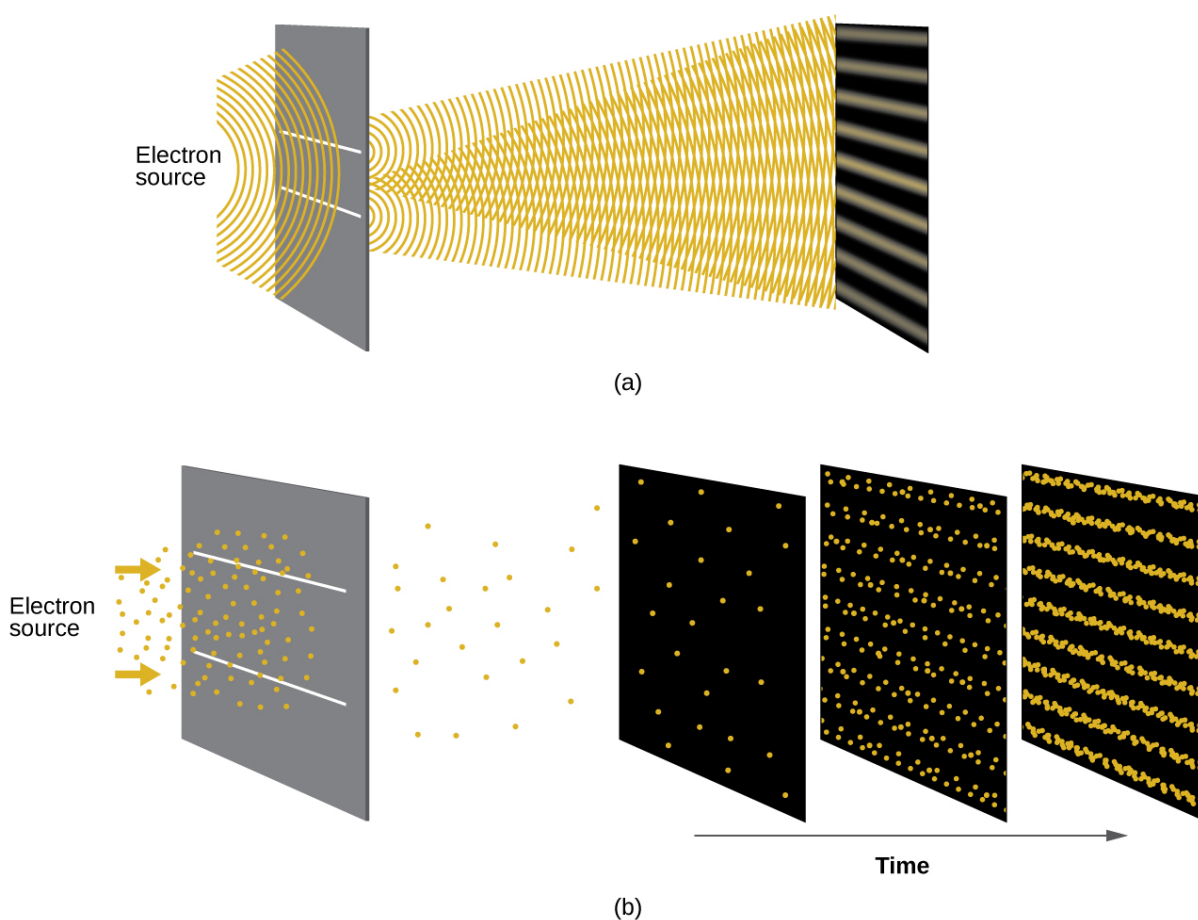
For a circular orbit of radius r , the circumference is $2\pi r$, and so de Broglie's condition is:

Equation:

$$2\pi r = n\lambda, \quad n = 1, 2, 3, \dots$$

Shortly after de Broglie proposed the wave nature of matter, two scientists at Bell Laboratories, C. J. Davisson and L. H. Germer, demonstrated experimentally that electrons can exhibit wavelike behavior by showing an interference pattern for electrons travelling through a regular atomic pattern in a crystal. The regularly spaced atomic layers served as slits, as used in other interference experiments. Since the spacing between the layers serving as slits needs to be similar in size to the wavelength of the tested wave for an interference pattern to form, Davisson and Germer used a crystalline nickel target for their "slits," since the spacing of the atoms within the lattice was approximately the same as the de Broglie wavelengths of the electrons that they used. [\[link\]](#) shows an interference pattern. It is strikingly similar

to the interference patterns for light shown in Electromagnetic Energy for light passing through two closely spaced, narrow slits. The wave–particle duality of matter can be seen in [\[link\]](#) by observing what happens if electron collisions are recorded over a long period of time. Initially, when only a few electrons have been recorded, they show clear particle-like behavior, having arrived in small localized packets that appear to be random. As more and more electrons arrived and were recorded, a clear interference pattern that is the hallmark of wavelike behavior emerged. Thus, it appears that while electrons are small localized particles, their motion does not follow the equations of motion implied by classical mechanics, but instead it is governed by some type of a wave equation. Thus the wave–particle duality first observed with photons is actually a fundamental behavior intrinsic to all quantum particles.



(a) The interference pattern for electrons passing through very closely spaced slits demonstrates that quantum particles such as electrons can exhibit wavelike

behavior. (b) The experimental results illustrated here demonstrate the wave–particle duality in electrons.

Note:

View the Dr. Quantum – Double Slit Experiment [cartoon](#) for an easy-to-understand description of wave–particle duality and the associated experiments.

Example:

Calculating the Wavelength of a Particle

If an electron travels at a velocity of $1.000 \times 10^7 \text{ m s}^{-1}$ and has a mass of $9.109 \times 10^{-28} \text{ g}$, what is its wavelength?

Solution

We can use de Broglie’s equation to solve this problem, but we first must do a unit conversion of Planck’s constant. You learned earlier that $1 \text{ J} = 1 \text{ kg m}^2/\text{s}^2$. Thus, we can write $h = 6.626 \times 10^{-34} \text{ J s}$ as $6.626 \times 10^{-34} \text{ kg m}^2/\text{s}$.

Equation:

$$\lambda = \frac{h}{mv}$$

Equation:

$$\begin{aligned} &= \frac{6.626 \times 10^{-34} \text{ kg m}^2/\text{s}}{(9.109 \times 10^{-31} \text{ kg})(1.000 \times 10^7 \text{ m/s})} \\ &= 7.274 \times 10^{-11} \text{ m} \end{aligned}$$

This is a small value, but it is significantly larger than the size of an electron in the classical (particle) view. This size is the same order of magnitude as the size of an atom. This means that electron wavelike behavior is going to be noticeable in an atom.

Check Your Learning

Calculate the wavelength of a softball with a mass of 100 g traveling at a velocity of 35 m s^{-1} , assuming that it can be modeled as a single particle.

Note:

Answer:

$$1.9 \times 10^{-34} \text{ m.}$$

We never think of a thrown softball having a wavelength, since this wavelength is so small it is impossible for our senses or any known instrument to detect (strictly speaking, the wavelength of a real baseball would correspond to the wavelengths of its constituent atoms and molecules, which, while much larger than this value, would still be microscopically tiny). The de Broglie wavelength is only appreciable for matter that has a very small mass and/or a very high velocity.

Werner Heisenberg considered the limits of how accurately we can measure properties of an electron or other microscopic particles. He determined that there is a fundamental limit to how accurately one can measure both a particle's position and its momentum simultaneously. The more accurately we measure the momentum of a particle, the less accurately we can determine its position at that time, and vice versa. This is summed up in what we now call the **Heisenberg uncertainty principle**: *It is fundamentally impossible to determine simultaneously and exactly both the momentum and the position of a particle.* For a particle of mass m moving with velocity v_x in the x direction (or equivalently with momentum p_x), the product of the uncertainty in the position, Δx , and the uncertainty in the momentum, Δp_x , must be greater than or equal to $\frac{\hbar}{2}$ (where $\hbar = \frac{h}{2\pi}$, the value of Planck's constant divided by 2π).

Equation:

$$\Delta x \times \Delta p_x = (\Delta x) (m\Delta v) \geq \frac{\hbar}{2}$$

This equation allows us to calculate the limit to how precisely we can know both the simultaneous position of an object and its momentum. For example, if we improve our measurement of an electron's position so that the uncertainty in the position (Δx) has a value of, say, 1 pm (10^{-12} m, about 1% of the diameter of a hydrogen atom), then our determination of its momentum must have an uncertainty with a value of at least

Equation:

$$\left[\Delta p = m\Delta v = \frac{h}{(2\Delta x)} \right] = \frac{(1.055 \times 10^{-34} \text{ kg m}^2/\text{s})}{(2 \times 1 \times 10^{-12} \text{ m})} = 5 \times 10^{-23} \text{ kg m/s}.$$

The value of $\hbar$ is not large, so the uncertainty in the position or momentum of a macroscopic object like a baseball is too insignificant to observe. However, the mass of a microscopic object such as an electron is small enough that the uncertainty can be large and significant.

It should be noted that Heisenberg's uncertainty principle is not just limited to uncertainties in position and momentum, but it also links other dynamical variables. For example, when an atom absorbs a photon and makes a transition from one energy state to another, the uncertainty in the energy and the uncertainty in the time required for the transition are similarly related, as $\Delta E \Delta t \geq \frac{\hbar}{2}$.

Heisenberg's principle imposes ultimate limits on what is knowable in science. The uncertainty principle can be shown to be a consequence of wave-particle duality, which lies at the heart of what distinguishes modern quantum theory from classical mechanics.

Note:

Read this [article](#) that describes a recent macroscopic demonstration of the uncertainty principle applied to microscopic objects.

The Quantum-Mechanical Model of an Atom

Shortly after de Broglie published his ideas that the electron in a hydrogen atom could be better thought of as being a circular standing wave instead of a particle moving in quantized circular orbits, Erwin Schrödinger extended de Broglie's work by deriving what is today known as the Schrödinger equation. When Schrödinger applied his equation to hydrogen-like atoms, he was able to reproduce Bohr's expression for the energy and, thus, the Rydberg formula governing hydrogen spectra. Schrödinger described electrons as three-dimensional stationary waves, or **wavefunctions**, represented by the Greek letter psi, ψ . A few years later, Max Born proposed an interpretation of the wavefunction ψ that is still accepted today: Electrons are still particles, and so the waves represented by ψ are not physical waves but, instead, are complex probability amplitudes. The square of the magnitude of a wavefunction $|\psi|^2$ describes the probability of the quantum particle being

present near a certain location in space. This means that wavefunctions can be used to determine the distribution of the electron's density with respect to the nucleus in an atom. In the most general form, the Schrödinger equation can be written as:

Equation:

$$\hat{H}\psi = E\psi$$

$\hat{H}$ is the Hamiltonian operator, a set of mathematical operations representing the total energy of the quantum particle (such as an electron in an atom), ψ is the wavefunction of this particle that can be used to find the special distribution of the probability of finding the particle, and E is the actual value of the total energy of the particle.

Schrödinger's work, as well as that of Heisenberg and many other scientists following in their footsteps, is generally referred to as **quantum mechanics**.

Note:

You may also have heard of Schrödinger because of his famous thought experiment. [This story](#) explains the concepts of superposition and entanglement as related to a cat in a box with poison.

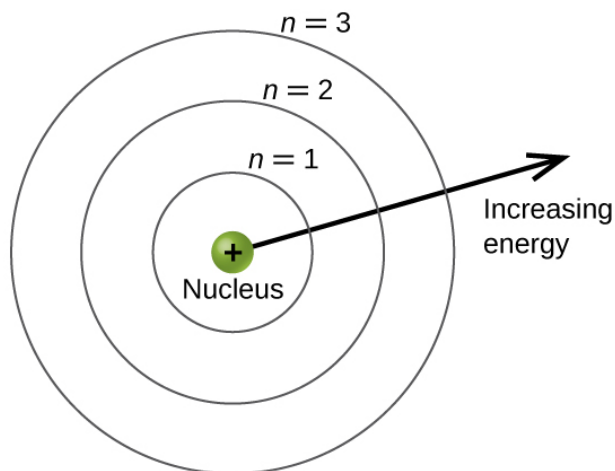
Understanding Quantum Theory of Electrons in Atoms

The goal of this section is to understand the electron orbitals (location of electrons in atoms), their different energies, and other properties. The use of quantum theory provides the best understanding to these topics. This knowledge is a precursor to chemical bonding.

As was described previously, electrons in atoms can exist only on discrete energy levels but not between them. It is said that the energy of an electron in an atom is quantized, that is, it can be equal only to certain specific values and can jump from one energy level to another but not transition smoothly or stay between these levels.

The energy levels are labeled with an n value, where $n = 1, 2, 3, \dots$. Generally speaking, the energy of an electron in an atom is greater for greater values of n . This number, n , is referred to as the principal quantum number. The **principal quantum number** defines the location of the energy level. It is essentially the same concept as

the n in the Bohr atom description. Another name for the principal quantum number is the shell number. The **shells** of an atom can be thought of concentric circles radiating out from the nucleus. The electrons that belong to a specific shell are most likely to be found within the corresponding circular area. The further we proceed from the nucleus, the higher the shell number, and so the higher the energy level ([link](#)). The positively charged protons in the nucleus stabilize the electronic orbitals by electrostatic attraction between the positive charges of the protons and the negative charges of the electrons. So the further away the electron is from the nucleus, the greater the energy it has.



Different shells are numbered by principal quantum numbers.

This quantum mechanical model for where electrons reside in an atom can be used to look at electronic transitions, the events when an electron moves from one energy level to another. If the transition is to a higher energy level, energy is absorbed, and the energy change has a positive value. To obtain the amount of energy necessary for the transition to a higher energy level, a photon is absorbed by the atom. A transition to a lower energy level involves a release of energy, and the energy change is negative. This process is accompanied by emission of a photon by the atom. The following equation summarizes these relationships and is based on the hydrogen atom:

Equation:

$$\begin{aligned}\Delta E &= E_{\text{final}} - E_{\text{initial}} \\ &= -2.18 \times 10^{-18} \left(\frac{1}{n_f^2} - \frac{1}{n_i^2} \right) \text{ J}\end{aligned}$$

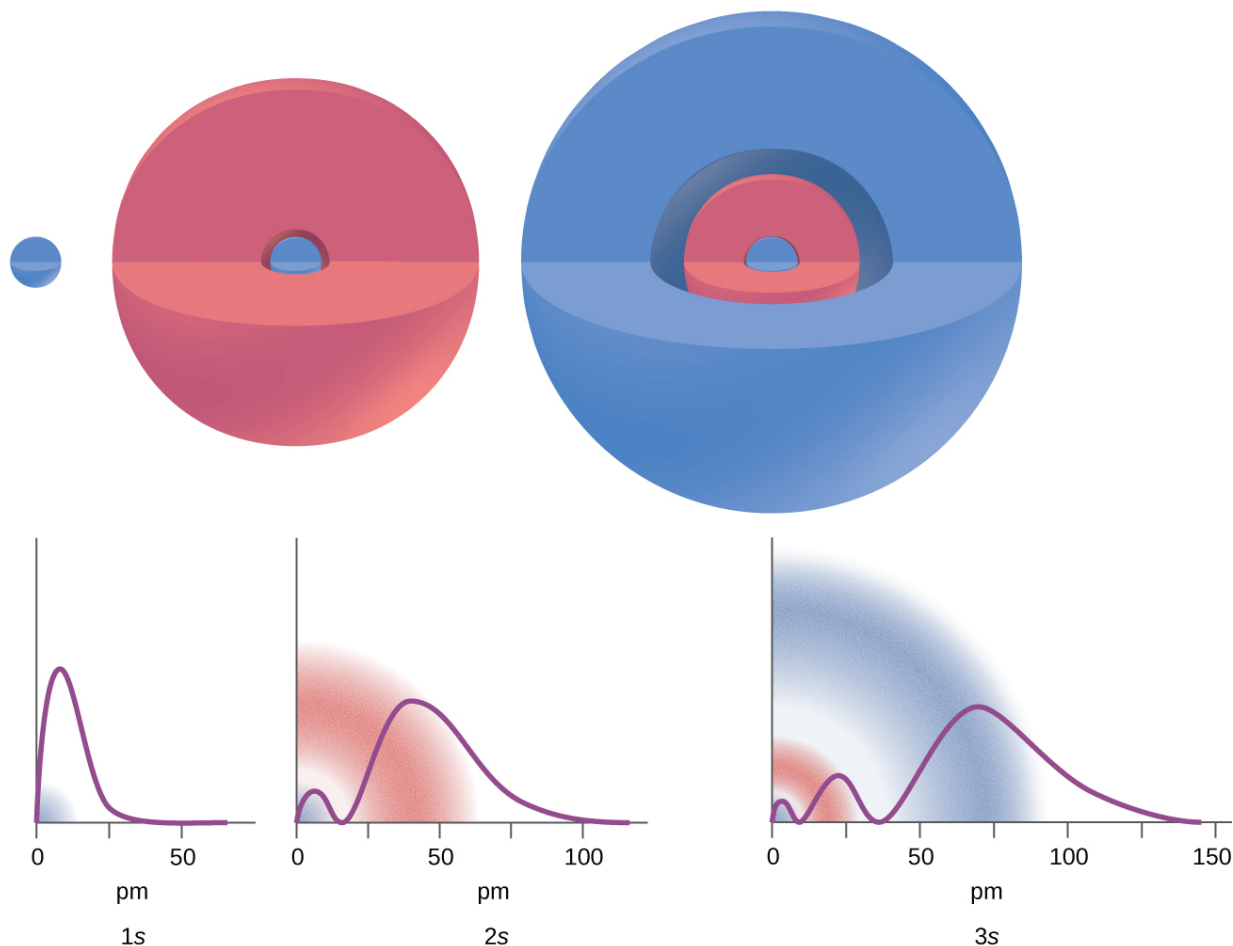
The values n_f and n_i are the final and initial energy states of the electron. [\[link\]](#) in the previous section of the chapter demonstrates calculations of such energy changes.

The principal quantum number is one of three quantum numbers used to characterize an orbital. An **atomic orbital** is a general region in an atom within which an electron is most probable to reside. The quantum mechanical model specifies the probability of finding an electron in the three-dimensional space around the nucleus and is based on solutions of the Schrödinger equation. In addition, the principal quantum number defines the energy of an electron in a hydrogen or hydrogen-like atom or an ion (an atom or an ion with only one electron) and the general region in which discrete energy levels of electrons in a multi-electron atoms and ions are located.

Another quantum number is l , the **secondary (angular momentum) quantum number**. It is an integer that may take the values, $l = 0, 1, 2, \dots, n - 1$. This means that an orbital with $n = 1$ can have only one value of l , $l = 0$, whereas $n = 2$ permits $l = 0$ and $l = 1$, and so on. Whereas the principal quantum number, n , defines the general size and energy of the orbital, the secondary quantum number l specifies the shape of the orbital. Orbitals with the same value of l define a **subshell**.

Orbitals with $l = 0$ are called **s orbitals** and they make up the s subshells. The value $l = 1$ corresponds to the *p* orbitals. For a given n , **p orbitals** constitute a *p* subshell (e.g., $3p$ if $n = 3$). The orbitals with $l = 2$ are called the **d orbitals**, followed by the *f*-, *g*-, and *h*-orbitals for $l = 3, 4$, and 5 .

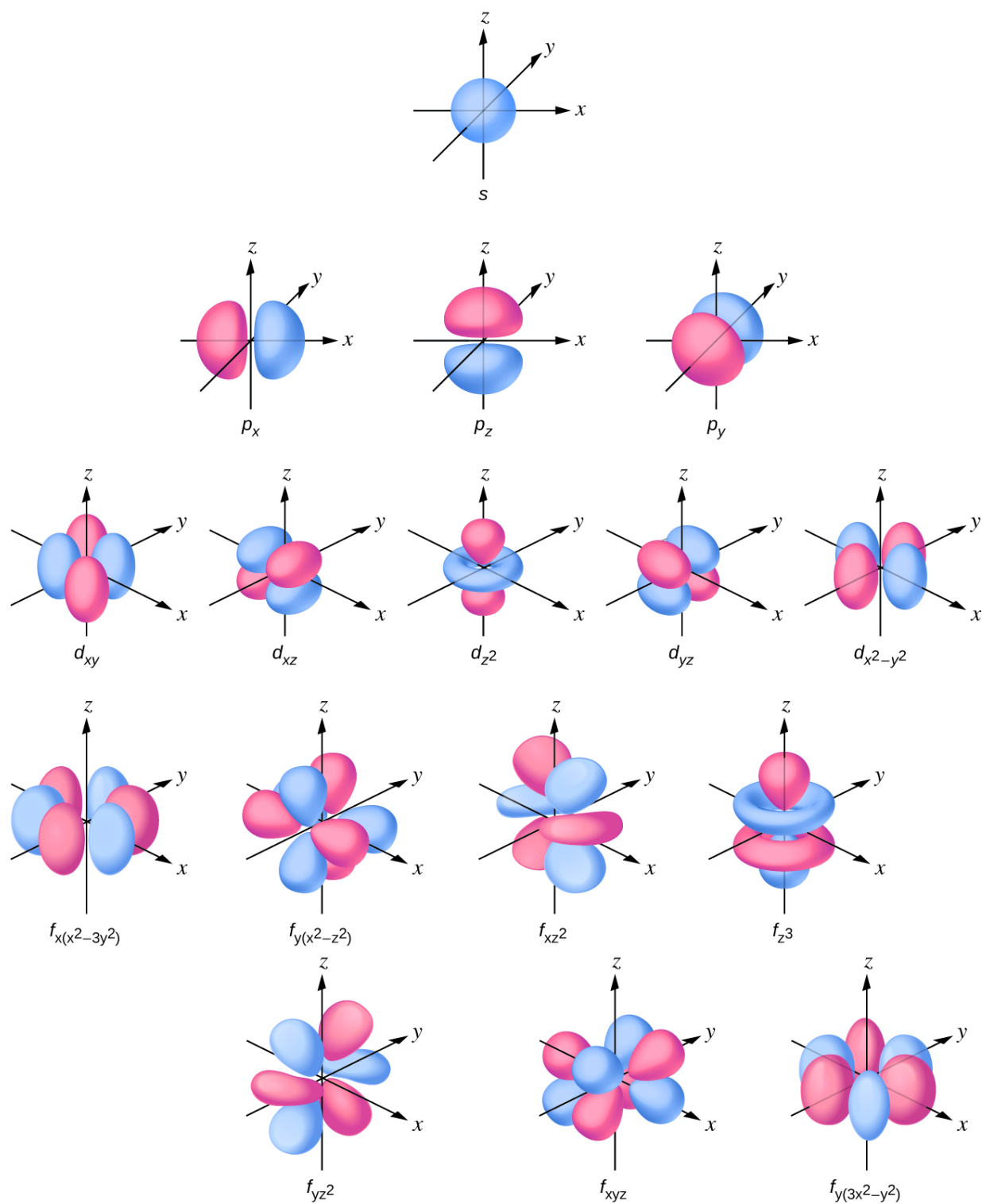
There are certain distances from the nucleus at which the probability density of finding an electron located at a particular orbital is zero. In other words, the value of the wavefunction ψ is zero at this distance for this orbital. Such a value of radius r is called a radial node. The number of radial nodes in an orbital is $n - l - 1$.



The graphs show the probability (y axis) of finding an electron for the 1s, 2s, 3s orbitals as a function of distance from the nucleus.

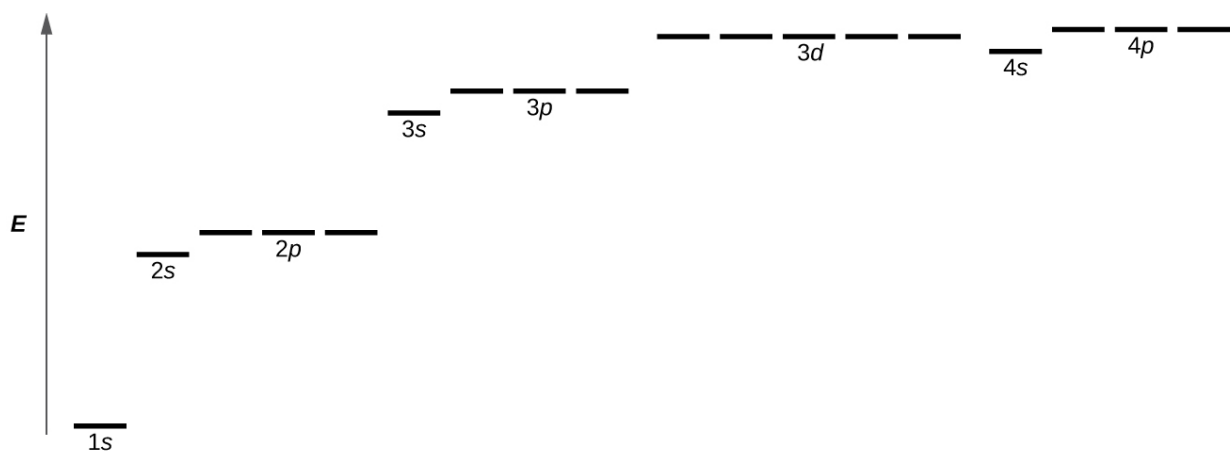
Consider the examples in [\[link\]](#). The orbitals depicted are of the s type, thus $l = 0$ for all of them. It can be seen from the graphs of the probability densities that there are $1 - 0 - 1 = 0$ places where the density is zero (nodes) for 1s ($n = 1$), $2 - 0 - 1 = 1$ node for 2s, and $3 - 0 - 1 = 2$ nodes for the 3s orbitals.

The s subshell electron density distribution is spherical and the *p* subshell has a dumbbell shape. The *d* and *f* orbitals are more complex. These shapes represent the three-dimensional regions within which the electron is likely to be found.



Shapes of s, p, d, and f orbitals.

The **magnetic quantum number**, m_l , specifies the relative spatial orientation of a particular orbital. Generally speaking, m_l can be equal to $-l, -(l-1), \dots, 0, \dots, (l-1), l$. The total number of possible orbitals with the same value of l (that is, in the same subshell) is $2l + 1$. Thus, there is one s -orbital in an s subshell ($l = 0$), there are three p -orbitals in a p subshell ($l = 1$), five d -orbitals in a d subshell ($l = 2$), seven f -orbitals in an f subshell ($l = 3$), and so forth. The principal quantum number defines the general value of the electronic energy. The angular momentum quantum number determines the shape of the orbital. And the magnetic quantum number specifies orientation of the orbital in space, as can be seen in [\[link\]](#).



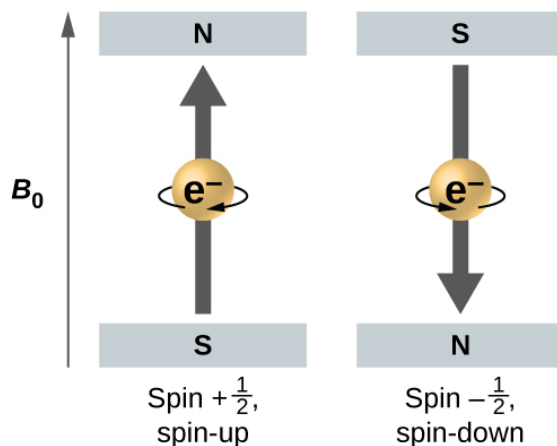
The chart shows the energies of electron orbitals in a multi-electron atom.

[\[link\]](#) illustrates the energy levels for various orbitals. The number before the orbital name (such as $2s$, $3p$, and so forth) stands for the principal quantum number, n . The letter in the orbital name defines the subshell with a specific angular momentum quantum number $l = 0$ for s orbitals, 1 for p orbitals, 2 for d orbitals. Finally, there are more than one possible orbitals for $l \geq 1$, each corresponding to a specific value of m_l . In the case of a hydrogen atom or a one-electron ion (such as He^+ , Li^{2+} , and so on), energies of all the orbitals with the same n are the same. This is called a degeneracy, and the energy levels for the same principal quantum number, n , are called **degenerate orbitals**. However, in atoms with more than one electron, this degeneracy is eliminated by the electron–electron interactions, and orbitals that belong to different subshells have different energies, as shown on [\[link\]](#). Orbitals within the same subshell are still degenerate and have the same energy.

While the three quantum numbers discussed in the previous paragraphs work well for describing electron orbitals, some experiments showed that they were not sufficient to explain all observed results. It was demonstrated in the 1920s that when hydrogen-line spectra are examined at extremely high resolution, some lines are actually not single peaks but, rather, pairs of closely spaced lines. This is the so-called fine structure of the spectrum, and it implies that there are additional small differences in energies of electrons even when they are located in the same orbital. These observations led Samuel Goudsmit and George Uhlenbeck to propose that electrons have a fourth quantum number. They called this the **spin quantum number**, or m_s .

The other three quantum numbers, n , l , and m_l , are properties of specific atomic orbitals that also define in what part of the space an electron is most likely to be located. Orbitals are a result of solving the Schrödinger equation for electrons in atoms. The electron spin is a different kind of property. It is a completely quantum phenomenon with no analogues in the classical realm. In addition, it cannot be derived from solving the Schrödinger equation and is not related to the normal spatial coordinates (such as the Cartesian x , y , and z). Electron spin describes an intrinsic electron "rotation" or "spinning." Each electron acts as a tiny magnet or a tiny rotating object with an angular momentum, or as a loop with an electric current, even though this rotation or current cannot be observed in terms of spatial coordinates.

The magnitude of the overall electron spin can only have one value, and an electron can only "spin" in one of two quantized states. One is termed the α state, with the z component of the spin being in the positive direction of the z axis. This corresponds to the spin quantum number $m_s = \frac{1}{2}$. The other is called the β state, with the z component of the spin being negative and $m_s = -\frac{1}{2}$. Any electron, regardless of the atomic orbital it is located in, can only have one of those two values of the spin quantum number. The energies of electrons having $m_s = -\frac{1}{2}$ and $m_s = \frac{1}{2}$ are different if an external magnetic field is applied.



Electrons with spin values $\pm \frac{1}{2}$ in an external magnetic field.

[\[link\]](#) illustrates this phenomenon. An electron acts like a tiny magnet. Its moment is directed up (in the positive direction of the z axis) for the $\frac{1}{2}$ spin quantum number and down (in the negative z direction) for the spin quantum number of $-\frac{1}{2}$. A magnet has a lower energy if its magnetic moment is aligned with the external magnetic field (the left electron on [\[link\]](#)) and a higher energy for the magnetic moment being opposite to the applied field. This is why an electron with $m_s = \frac{1}{2}$ has a slightly lower energy in an external field in the positive z direction, and an electron with $m_s = -\frac{1}{2}$ has a slightly higher energy in the same field. This is true even for an electron occupying the same orbital in an atom. A spectral line corresponding to a transition for electrons from the same orbital but with different spin quantum numbers has two possible values of energy; thus, the line in the spectrum will show a fine structure splitting.

The Pauli Exclusion Principle

An electron in an atom is completely described by four quantum numbers: n , l , m_l , and m_s . The first three quantum numbers define the orbital and the fourth quantum number describes the intrinsic electron property called spin. An Austrian physicist Wolfgang Pauli formulated a general principle that gives the last piece of information that we need to understand the general behavior of electrons in atoms. The **Pauli exclusion principle** can be formulated as follows: No two electrons in the same atom can have exactly the same set of all the four quantum numbers. What this means is that two electrons can share the same orbital (the same set of the quantum numbers n , l , and m_l) only if their spin quantum numbers m_s have different values. Since the spin quantum number can only have two values ($\pm \frac{1}{2}$), no more than two electrons can occupy the same orbital (and if two electrons are located in the same orbital, they must have opposite spins). Therefore, any atomic orbital can be populated by only zero, one, or two electrons.

The properties and meaning of the quantum numbers of electrons in atoms are briefly summarized in [\[link\]](#).

Quantum Numbers, Their Properties, and Significance			
Name	Symbol	Allowed values	Physical meaning
principal quantum number	n	1, 2, 3, 4, ...	shell, the general region for the value of energy for an electron on the orbital
angular momentum or azimuthal quantum number	l	$0 \leq l \leq n - 1$	subshell, the shape of the orbital
magnetic quantum number	m_l	$-l \leq m_l \leq l$	orientation of the orbital
spin quantum number	m_s	$\frac{1}{2}, -\frac{1}{2}$	direction of the intrinsic quantum “spinning” of the electron

Example:

Working with Shells and Subshells

Indicate the number of subshells, the number of orbitals in each subshell, and the values of l and m_l for the orbitals in the $n = 4$ shell of an atom.

Solution

For $n = 4$, l can have values of 0, 1, 2, and 3. Thus, s , p , d , and f subshells are found in the $n = 4$ shell of an atom. For $l = 0$ (the s subshell), m_l can only be 0. Thus, there is only one $4s$ orbital. For $l = 1$ (p -type orbitals), m can have values of $-1, 0, +1$, so we find three $4p$ orbitals. For $l = 2$ (d -type orbitals), m_l can have values of $-2, -1, 0, +1, +2$, so we have five $4d$ orbitals. When $l = 3$ (f -type orbitals), m_l can have values of $-3, -2, -1, 0, +1, +2, +3$, and we can have seven $4f$ orbitals. Thus, we find a total of 16 orbitals in the $n = 4$ shell of an atom.

Check Your Learning

Identify the subshell in which electrons with the following quantum numbers are found: (a) $n = 3, l = 1$; (b) $n = 5, l = 3$; (c) $n = 2, l = 0$.

Note:

Answer:

(a) $3p$ (b) $5f$ (c) $2s$

Example:

Maximum Number of Electrons

Calculate the maximum number of electrons that can occupy a shell with (a) $n = 2$, (b) $n = 5$, and (c) n as a variable. Note you are only looking at the orbitals with the specified n value, not those at lower energies.

Solution

(a) When $n = 2$, there are four orbitals (a single $2s$ orbital, and three orbitals labeled $2p$). These four orbitals can contain eight electrons.

(b) When $n = 5$, there are five subshells of orbitals that we need to sum:

Equation:

$$\begin{array}{r} 1 \text{ orbital labeled } 5s \\ 3 \text{ orbitals labeled } 5p \\ 5 \text{ orbitals labeled } 5d \\ 7 \text{ orbitals labeled } 5f \\ \hline +9 \text{ orbitals labeled } 5g \\ \hline 25 \text{ orbitals total} \end{array}$$

Again, each orbital holds two electrons, so 50 electrons can fit in this shell.

(c) The number of orbitals in any shell n will equal n^2 . There can be up to two electrons in each orbital, so the maximum number of electrons will be $2 \times n^2$.

Check Your Learning

If a shell contains a maximum of 32 electrons, what is the principal quantum number, n ?

Note:

Answer:

$$n = 4$$

Example:**Working with Quantum Numbers**

Complete the following table for atomic orbitals:

Orbital	n	l	m_l degeneracy	Radial nodes (no.)
$4f$				
	4	1		
	7		7	3
$5d$				

Solution

The table can be completed using the following rules:

- The orbital designation is nl , where $l = 0, 1, 2, 3, 4, 5, \dots$ is mapped to the letter sequence $s, p, d, f, g, h, \dots$,
- The m_l degeneracy is the number of orbitals within an l subshell, and so is $2l + 1$ (there is one s orbital, three p orbitals, five d orbitals, seven f orbitals, and so forth).
- The number of radial nodes is equal to $n - l - 1$.

Orbital	n	l	m_l degeneracy	Radial nodes (no.)
$4f$	4	3	7	0
$4p$	4	1	3	2
$7f$	7	3	7	3
$5d$	5	2	5	2

Check Your Learning

How many orbitals have $l = 2$ and $n = 3$?

Note:

Answer:

The five degenerate $3d$ orbitals

Key Concepts and Summary

Macroscopic objects act as particles. Microscopic objects (such as electrons) have properties of both a particle and a wave. Their exact trajectories cannot be determined. The quantum mechanical model of atoms describes the three-dimensional position of the electron in a *probabilistic* manner according to a mathematical function called a wavefunction, often denoted as ψ . Atomic wavefunctions are also called orbitals. The squared magnitude of the wavefunction describes the distribution of the probability of finding the electron in a particular region in space. Therefore, atomic orbitals describe the areas in an atom where electrons are most likely to be found.

An atomic orbital is characterized by three quantum numbers. The principal quantum number, n , can be any positive integer. The general region for value of energy of the orbital and the average distance of an electron from the nucleus are related to n . Orbitals having the same value of n are said to be in the same shell. The secondary (angular momentum) quantum number, l , can have any integer value from 0 to $n - 1$.

This quantum number describes the shape or type of the orbital. Orbitals with the same principal quantum number and the same l value belong to the same subshell. The magnetic quantum number, m_l , with $2l + 1$ values ranging from $-l$ to $+l$, describes the orientation of the orbital in space. In addition, each electron has a spin quantum number, m_s , that can be equal to $\pm \frac{1}{2}$. No two electrons in the same atom can have the same set of values for all the four quantum numbers.

Chemistry End of Chapter Exercises

Exercise:

Problem:

How are the Bohr model and the quantum mechanical model of the hydrogen atom similar? How are they different?

Solution:

Both models have a central positively charged nucleus with electrons moving about the nucleus in accordance with the Coulomb electrostatic potential. The Bohr model *assumes* that the electrons move in circular orbits that have quantized energies, angular momentum, and radii that are specified by a single quantum number, $n = 1, 2, 3, \dots$, but this quantization is an ad hoc assumption made by Bohr to incorporate quantization into an essentially classical mechanics description of the atom. Bohr also assumed that electrons orbiting the nucleus normally do not emit or absorb electromagnetic radiation, but do so when the electron switches to a different orbit. In the quantum mechanical model, the electrons do not move in precise orbits (such orbits violate the Heisenberg uncertainty principle) and, instead, a probabilistic interpretation of the electron's position at any given instant is used, with a mathematical function ψ called a wavefunction that can be used to determine the electron's spatial probability distribution. These wavefunctions, or orbitals, are three-dimensional stationary waves that can be specified by three quantum numbers that arise naturally from their underlying mathematics (no ad hoc assumptions required): the principal quantum number, n (the same one used by Bohr), which specifies shells such that orbitals having the same n all have the same energy and approximately the same spatial extent; the angular momentum quantum number l , which is a measure of the orbital's angular momentum and corresponds to the orbitals' general shapes, as well as specifying subshells such that orbitals having the same l (and n) all have the same energy; and the orientation quantum number m , which is a measure of the z component of the angular momentum and corresponds to the orientations of the orbitals. The Bohr model gives the

same expression for the energy as the quantum mechanical expression and, hence, both properly account for hydrogen's discrete spectrum (an example of getting the right answers for the wrong reasons, something that many chemistry students can sympathize with), but gives the wrong expression for the angular momentum (Bohr orbits necessarily all have non-zero angular momentum, but some quantum orbitals [s orbitals] can have zero angular momentum).

Exercise:

Problem:

What are the allowed values for each of the four quantum numbers: n , l , m_l , and m_s ?

Exercise:

Problem:

Describe the properties of an electron associated with each of the following four quantum numbers: n , l , m_l , and m_s .

Solution:

n determines the general range for the value of energy and the probable distances that the electron can be from the nucleus. l determines the shape of the orbital. m_l determines the orientation of the orbitals of the same l value with respect to one another. m_s determines the spin of an electron.

Exercise:

Problem: Answer the following questions:

(a) Without using quantum numbers, describe the differences between the shells, subshells, and orbitals of an atom.

(b) How do the quantum numbers of the shells, subshells, and orbitals of an atom differ?

Exercise:

Problem:

Identify the subshell in which electrons with the following quantum numbers are found:

(a) $n = 2$, $l = 1$

(b) $n = 4, l = 2$

(c) $n = 6, l = 0$

Solution:

(a) $2p$; (b) $4d$; (c) $6s$

Exercise:

Problem:

Which of the subshells described in the previous question contain degenerate orbitals? How many degenerate orbitals are in each?

Exercise:

Problem:

Identify the subshell in which electrons with the following quantum numbers are found:

(a) $n = 3, l = 2$

(b) $n = 1, l = 0$

(c) $n = 4, l = 3$

Solution:

(a) $3d$; (b) $1s$; (c) $4f$

Exercise:

Problem:

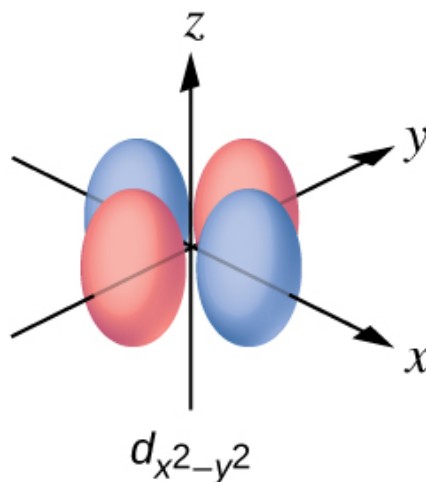
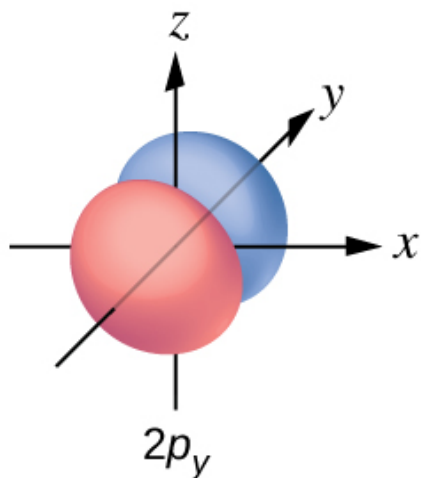
Which of the subshells described in the previous question contain degenerate orbitals? How many degenerate orbitals are in each?

Exercise:

Problem:

Sketch the boundary surface of a $d_{x^2-y^2}$ and a p_y orbital. Be sure to show and label the axes.

Solution:



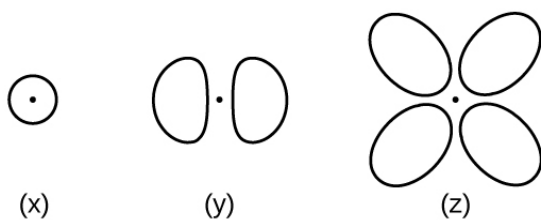
Exercise:

Problem:

Sketch the p_x and d_{xz} orbitals. Be sure to show and label the coordinates.

Exercise:

Problem: Consider the orbitals shown here in outline.



(a) What is the maximum number of electrons contained in an orbital of type (x)? Of type (y)? Of type (z)?

(b) How many orbitals of type (x) are found in a shell with $n = 2$? How many of type (y)? How many of type (z)?

(c) Write a set of quantum numbers for an electron in an orbital of type (x) in a shell with $n = 4$. Of an orbital of type (y) in a shell with $n = 2$. Of an orbital of type (z) in a shell with $n = 3$.

(d) What is the smallest possible n value for an orbital of type (x)? Of type (y)? Of type (z)?

(e) What are the possible l and m_l values for an orbital of type (x)? Of type (y)? Of type (z)?

Solution:

(a) x. 2, y. 2, z. 2; (b) x. 1, y. 3, z. 0; (c) x. $4\ 0\ 0\ \frac{1}{2}$, y. $2\ 1\ 0\ \frac{1}{2}$, z. $3\ 2\ 0\ \frac{1}{2}$; (d) x. 1, y. 2, z. 3; (e) x. $l = 0, m_l = 0$, y. $l = 1, m_l = -1, 0, \text{ or } +1$, z. $l = 2, m_l = -2, -1, 0, +1, +2$

Exercise:

Problem:

State the Heisenberg uncertainty principle. Describe briefly what the principle implies.

Exercise:

Problem:

How many electrons could be held in the second shell of an atom if the spin quantum number m_s could have three values instead of just two? (Hint: Consider the Pauli exclusion principle.)

Solution:

12

Exercise:

Problem:

Which of the following equations describe particle-like behavior? Which describe wavelike behavior? Do any involve both types of behavior? Describe the reasons for your choices.

(a) $c = \lambda\nu$

(b) $E = \frac{m\nu^2}{2}$

(c) $r = \frac{n^2 a_0}{Z}$

(d) $E = h\nu$

(e) $\lambda = \frac{h}{m\nu}$

Exercise:**Problem:**

Write a set of quantum numbers for each of the electrons with an n of 4 in a Se atom.

Solution:

n	l	m_l	s
4	0	0	$+\frac{1}{2}$
4	0	0	$-\frac{1}{2}$
4	1	-1	$+\frac{1}{2}$
4	1	0	$+\frac{1}{2}$
4	1	+1	$+\frac{1}{2}$
4	1	-1	$-\frac{1}{2}$

Glossary

secondary (angular momentum) quantum number (l)

quantum number distinguishing the different shapes of orbitals; it is also a measure of the orbital angular momentum

atomic orbital

mathematical function that describes the behavior of an electron in an atom (also called the wavefunction)

d orbital

region of space with high electron density that is either four lobed or contains a dumbbell and torus shape; describes orbitals with $l = 2$.

degenerate orbitals

orbitals that have the same energy

electron density

a measure of the probability of locating an electron in a particular region of space, it is equal to the squared absolute value of the wave function ψ

f orbital

multilobed region of space with high electron density, describes orbitals with $l = 3$

Heisenberg uncertainty principle

rule stating that it is impossible to exactly determine both certain conjugate dynamical properties such as the momentum and the position of a particle at the same time. The uncertainty principle is a consequence of quantum particles exhibiting wave-particle duality

magnetic quantum number (m_l)

quantum number signifying the orientation of an atomic orbital around the nucleus

p orbital

dumbbell-shaped region of space with high electron density, describes orbitals with $l = 1$

Pauli exclusion principle

specifies that no two electrons in an atom can have the same value for all four quantum numbers

principal quantum number (n)

quantum number specifying the shell an electron occupies in an atom

quantum mechanics

field of study that includes quantization of energy, wave-particle duality, and the Heisenberg uncertainty principle to describe matter

s orbital

spherical region of space with high electron density, describes orbitals with $l = 0$

shell

atomic orbitals with the same principal quantum number, n

spin quantum number (m_s)

number specifying the electron spin direction, either $+\frac{1}{2}$ or $-\frac{1}{2}$

subshell

atomic orbitals with the same values of n and l

wavefunction (ψ)

mathematical description of an atomic orbital that describes the shape of the orbital; it can be used to calculate the probability of finding the electron at any given location in the orbital, as well as dynamical variables such as the energy and the angular momentum

Electronic Structure of Atoms (Electron Configurations)

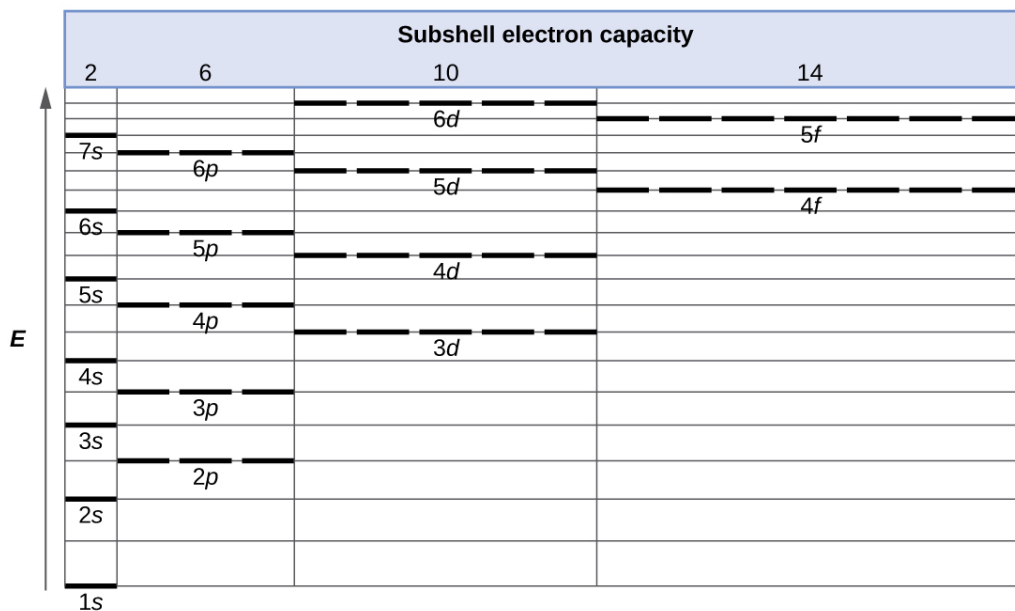
By the end of this section, you will be able to:

- Derive the predicted ground-state electron configurations of atoms
- Identify and explain exceptions to predicted electron configurations for atoms and ions
- Relate electron configurations to element classifications in the periodic table

Having introduced the basics of atomic structure and quantum mechanics, we can use our understanding of quantum numbers to determine how atomic orbitals relate to one another. This allows us to determine which orbitals are occupied by electrons in each atom. The specific arrangement of electrons in orbitals of an atom determines many of the chemical properties of that atom.

Orbital Energies and Atomic Structure

The energy of atomic orbitals increases as the principal quantum number, n , increases. In any atom with two or more electrons, the repulsion between the electrons makes energies of subshells with different values of l differ so that the energy of the orbitals increases within a shell in the order $s < p < d < f$. [\[link\]](#) depicts how these two trends in increasing energy relate. The $1s$ orbital at the bottom of the diagram is the orbital with electrons of lowest energy. The energy increases as we move up to the $2s$ and then $2p$, $3s$, and $3p$ orbitals, showing that the increasing n value has more influence on energy than the increasing l value for small atoms. However, this pattern does not hold for larger atoms. The $3d$ orbital is higher in energy than the $4s$ orbital. Such overlaps continue to occur frequently as we move up the chart.



Generalized energy-level diagram for atomic orbitals in an atom with two or more electrons (not to scale).

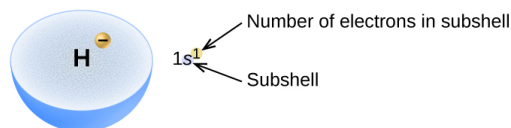
Electrons in successive atoms on the periodic table tend to fill low-energy orbitals first. Thus, many students find it confusing that, for example, the $5p$ orbitals fill immediately after the $4d$, and immediately before the $6s$. The filling order is based on observed experimental results, and has been confirmed by theoretical calculations. As the principal quantum number, n , increases, the size of the orbital increases and the electrons spend more time farther from the nucleus. Thus, the attraction to the nucleus is weaker and the energy associated with the orbital is higher (less stabilized). But this is not the only effect we have to take into account. Within each shell, as the value of l increases, the electrons are less penetrating (meaning there is less electron density found close to the nucleus), in the order $s > p > d > f$. Electrons that are closer to the nucleus slightly repel electrons that are farther out, offsetting the more dominant electron–nucleus attractions slightly (recall that all electrons have -1 charges, but nuclei have $+Z$ charges). This phenomenon is called shielding and will be discussed in more detail in the next section. Electrons in orbitals that experience more shielding are less stabilized and thus higher in energy. For small orbitals ($1s$

through $3p$), the increase in energy due to n is more significant than the increase due to l ; however, for larger orbitals the two trends are comparable and cannot be simply predicted. We will discuss methods for remembering the observed order.

The arrangement of electrons in the orbitals of an atom is called the **electron configuration** of the atom. We describe an electron configuration with a symbol that contains three pieces of information ([link](#)):

1. The number of the principal quantum shell, n ,
2. The letter that designates the orbital type (the subshell, l), and
3. A superscript number that designates the number of electrons in that particular subshell.

For example, the notation $2p^4$ (read "two-p-four") indicates four electrons in a p subshell ($l = 1$) with a principal quantum number (n) of 2. The notation $3d^8$ (read "three-d-eight") indicates eight electrons in the d subshell (i.e., $l = 2$) of the principal shell for which $n = 3$.

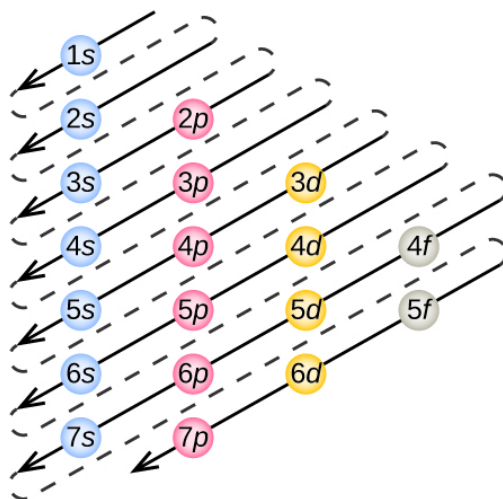


The diagram of an electron configuration specifies the subshell (n and l value, with letter symbol) and superscript number of electrons.

The Aufbau Principle

To determine the electron configuration for any particular atom, we can “build” the structures in the order of atomic numbers. Beginning with hydrogen, and continuing across the periods of the periodic table, we add one proton at a time to the nucleus and one electron to the proper subshell until we have described the electron configurations of all the elements. This procedure is called the **Aufbau principle**, from the German word *Aufbau*

(“to build up”). Each added electron occupies the subshell of lowest energy available (in the order shown in [\[link\]](#)), subject to the limitations imposed by the allowed quantum numbers according to the Pauli exclusion principle. Electrons enter higher-energy subshells only after lower-energy subshells have been filled to capacity. [\[link\]](#) illustrates the traditional way to remember the filling order for atomic orbitals. Since the arrangement of the periodic table is based on the electron configurations, [\[link\]](#) provides an alternative method for determining the electron configuration. The filling order simply begins at hydrogen and includes each subshell as you proceed in increasing Z order. For example, after filling the $3p$ block up to Ar, we see the orbital will be $4s$ (K, Ca), followed by the $3d$ orbitals.



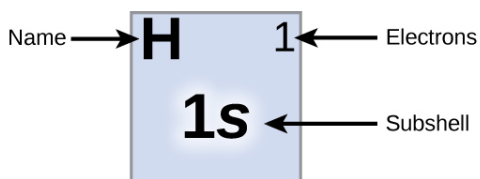
This diagram depicts the energy order for atomic orbitals and is useful for deriving ground-state electron configurations.

Period

Group

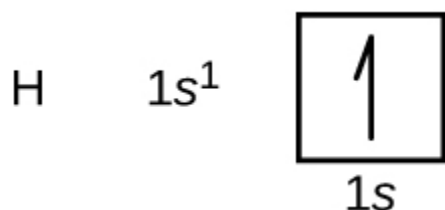
Electron Configuration Table

1	1	H	1																18
1		1s																	He
2	2	Li	Be											13	14	15	16	17	2
2		2s												B	C	N	O	F	Ne
3	3	Na	Mg																
3		3s												Al	Si	P	S	Cl	Ar
4	4	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn						
4		4s		3d										Ga	Ge	As	Se	Br	Kr
5	5	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd						
5		5s		4d										In	Sn	Sb	Te	I	Xe
6	6	Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg						
6		6s		5d										Tl	Pb	Bi	Po	At	Rn
7	7	Fr	Ra	Ac	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Nh	Fl	Mc	Lv	Ts	Og
7		7s		6d															
		Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu				
		4f																	
		Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr				
		5f																	

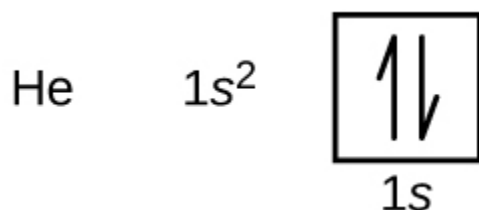


This periodic table shows the electron configuration for each subshell. By “building up” from hydrogen, this table can be used to determine the electron configuration for any atom on the periodic table.

We will now construct the ground-state electron configuration and orbital diagram for a selection of atoms in the first and second periods of the periodic table. **Orbital diagrams** are pictorial representations of the electron configuration, showing the individual orbitals and the pairing arrangement of electrons. We start with a single hydrogen atom (atomic number 1), which consists of one proton and one electron. Referring to [\[link\]](#) or [\[link\]](#), we would expect to find the electron in the 1s orbital. By convention, the $m_s = +\frac{1}{2}$ value is usually filled first. The electron configuration and the orbital diagram are:

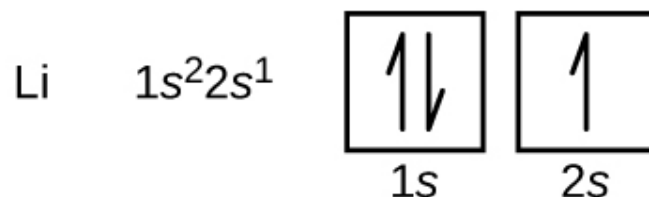


Following hydrogen is the noble gas helium, which has an atomic number of 2. The helium atom contains two protons and two electrons. The first electron has the same four quantum numbers as the hydrogen atom electron ($n = 1$, $l = 0$, $m_l = 0$, $m_s = +\frac{1}{2}$). The second electron also goes into the 1s orbital and fills that orbital. The second electron has the same n , l , and m_l quantum numbers, but must have the opposite spin quantum number, $m_s = -\frac{1}{2}$. This is in accord with the Pauli exclusion principle: No two electrons in the same atom can have the same set of four quantum numbers. For orbital diagrams, this means two arrows go in each box (representing two electrons in each orbital) and the arrows must point in opposite directions (representing paired spins). The electron configuration and orbital diagram of helium are:

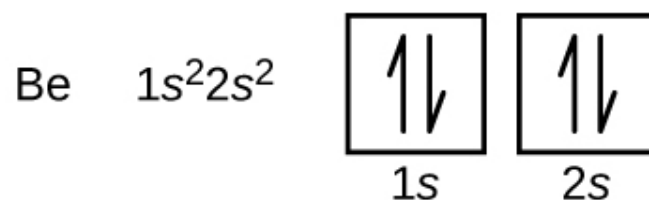


The $n = 1$ shell is completely filled in a helium atom.

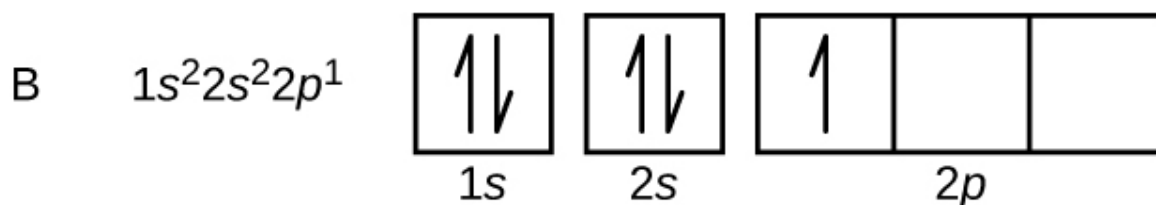
The next atom is the alkali metal lithium with an atomic number of 3. The first two electrons in lithium fill the 1s orbital and have the same sets of four quantum numbers as the two electrons in helium. The remaining electron must occupy the orbital of next lowest energy, the 2s orbital ([\[link\]](#) or [\[link\]](#)). Thus, the electron configuration and orbital diagram of lithium are:



An atom of the alkaline earth metal beryllium, with an atomic number of 4, contains four protons in the nucleus and four electrons surrounding the nucleus. The fourth electron fills the remaining space in the 2s orbital.

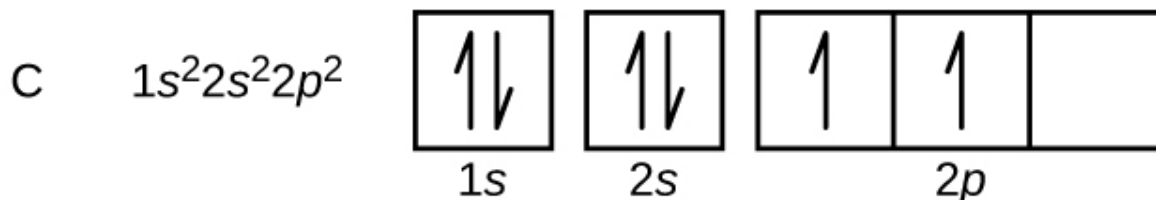


An atom of boron (atomic number 5) contains five electrons. The $n = 1$ shell is filled with two electrons and three electrons will occupy the $n = 2$ shell. Because any s subshell can contain only two electrons, the fifth electron must occupy the next energy level, which will be a 2p orbital. There are three degenerate 2p orbitals ($m_l = -1, 0, +1$) and the electron can occupy any one of these p orbitals. When drawing orbital diagrams, we include empty boxes to depict any empty orbitals in the same subshell that we are filling.

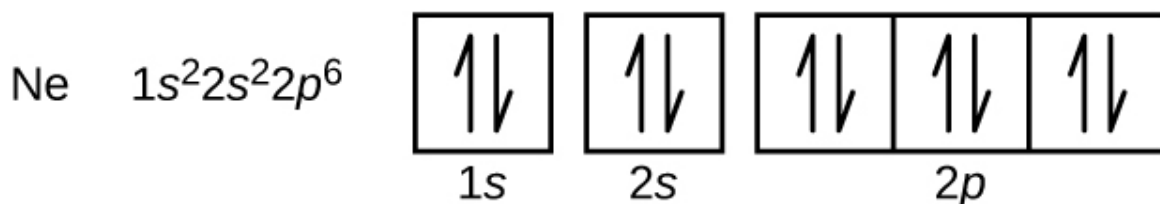
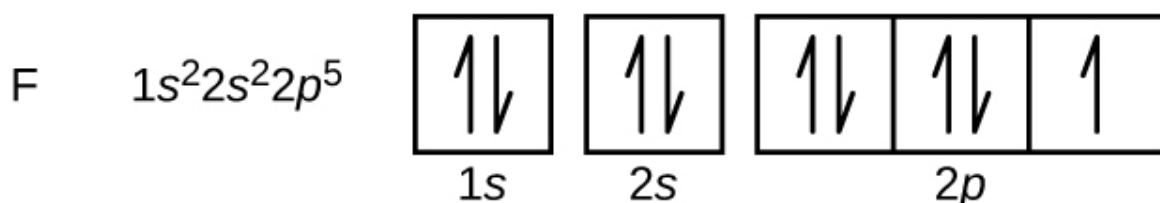
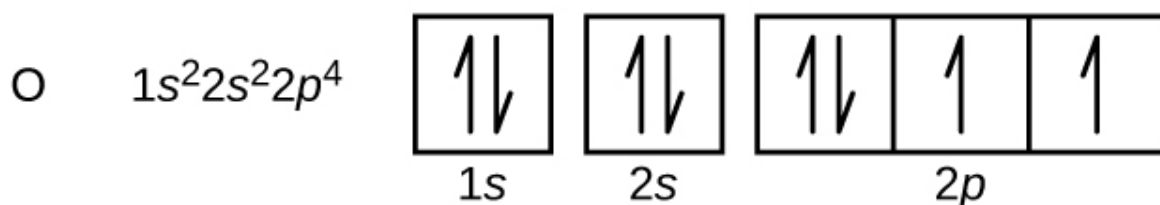
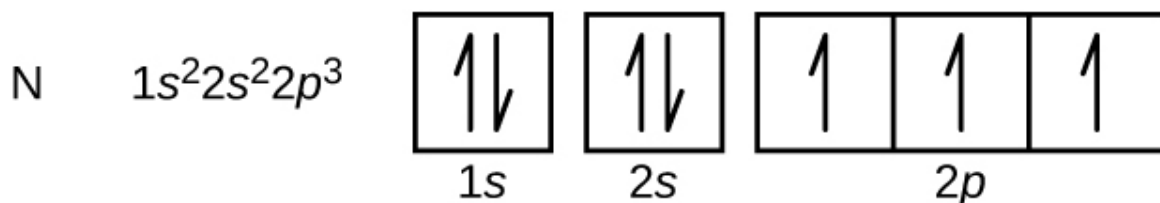


Carbon (atomic number 6) has six electrons. Four of them fill the 1s and 2s orbitals. The remaining two electrons occupy the 2p subshell. We now have a choice of filling one of the 2p orbitals and pairing the electrons or of leaving the electrons unpaired in two different, but degenerate, p orbitals. The orbitals are filled as described by **Hund's rule**: the lowest-energy configuration for an atom with electrons within a set of degenerate orbitals is that having the maximum number of unpaired electrons. Thus, the two electrons in the carbon 2p orbitals have identical n , l , and m_s quantum

numbers and differ in their m_l quantum number (in accord with the Pauli exclusion principle). The electron configuration and orbital diagram for carbon are:

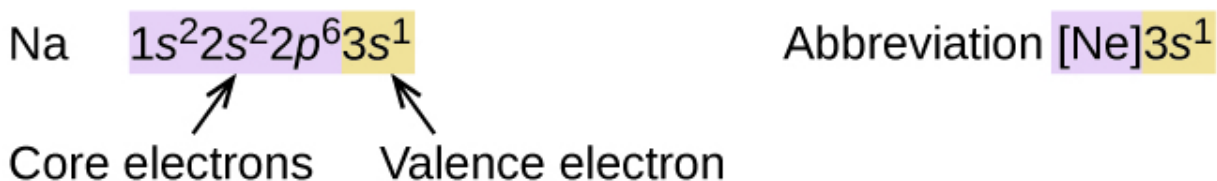


Nitrogen (atomic number 7) fills the $1s$ and $2s$ subshells and has one electron in each of the three $2p$ orbitals, in accordance with Hund's rule. These three electrons have unpaired spins. Oxygen (atomic number 8) has a pair of electrons in any one of the $2p$ orbitals (the electrons have opposite spins) and a single electron in each of the other two. Fluorine (atomic number 9) has only one $2p$ orbital containing an unpaired electron. All of the electrons in the noble gas neon (atomic number 10) are paired, and all of the orbitals in the $n = 1$ and the $n = 2$ shells are filled. The electron configurations and orbital diagrams of these four elements are:



The alkali metal sodium (atomic number 11) has one more electron than the neon atom. This electron must go into the lowest-energy subshell available, the 3s orbital, giving a $1s^2 2s^2 2p^6 3s^1$ configuration. The electrons occupying the outermost shell orbital(s) (highest value of n) are called **valence electrons**, and those occupying the inner shell orbitals are called **core electrons** ([\[link\]](#)). Since the core electron shells correspond to noble gas electron configurations, we can abbreviate electron configurations by writing the noble gas that matches the core electron configuration, along with the valence electrons in a condensed format. For our sodium example,

the symbol [Ne] represents core electrons, ($1s^2 2s^2 2p^6$) and our abbreviated or condensed configuration is $[\text{Ne}]3s^1$.



A core-abbreviated electron configuration (right) replaces the core electrons with the noble gas symbol whose configuration matches the core electron configuration of the other element.

Similarly, the abbreviated configuration of lithium can be represented as $[\text{He}]2s^1$, where [He] represents the configuration of the helium atom, which is identical to that of the filled inner shell of lithium. Writing the configurations in this way emphasizes the similarity of the configurations of lithium and sodium. Both atoms, which are in the alkali metal family, have only one electron in a valence s subshell outside a filled set of inner shells.

Equation:



The alkaline earth metal magnesium (atomic number 12), with its 12 electrons in a $[\text{Ne}]3s^2$ configuration, is analogous to its family member beryllium, $[\text{He}]2s^2$. Both atoms have a filled s subshell outside their filled inner shells. Aluminum (atomic number 13), with 13 electrons and the electron configuration $[\text{Ne}]3s^2 3p^1$, is analogous to its family member boron, $[\text{He}]2s^2 2p^1$.

The electron configurations of silicon (14 electrons), phosphorus (15 electrons), sulfur (16 electrons), chlorine (17 electrons), and argon (18

electrons) are analogous in the electron configurations of their outer shells to their corresponding family members carbon, nitrogen, oxygen, fluorine, and neon, respectively, except that the principal quantum number of the outer shell of the heavier elements has increased by one to $n = 3$. [\[link\]](#) shows the lowest energy, or ground-state, electron configuration for these elements as well as that for atoms of each of the known elements.

Period

Group

Electron Configuration Table

18

1																	18					
1	1	2															13	14	15	16	17	18
1	1	2															13	14	15	16	17	18
2	3	4															5	6	7	8	9	10
3	11	12															13	14	15	16	17	18
4	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36				
5	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54				
6	55	56	57	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86				
7	87	88	89	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118				

*

**

58	59	60	61	62	63	64	65	66	67	68	69	70	71
Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
6s ² 4f ²	6s ² 4f ³	6s ² 4f ⁴	6s ² 4f ⁵	6s ² 4f ⁶	6s ² 4f ⁷	6s ² 4f ⁷ 5d ¹	6s ² 4f ⁹	6s ² 4f ¹⁰	6s ² 4f ¹¹	6s ² 4f ¹²	6s ² 4f ¹³	6s ² 4f ¹⁴	6s ² 4f ¹⁴ 5d ¹
90	91	92	93	94	95	96	97	98	99	100	101	102	103
Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr
7s ² 6d ²	7s ² 5f ² 6d ¹	7s ² 5f ³ 6d ¹	7s ² 5f ⁴ 6d ¹	7s ² 5f ⁶	7s ² 5f ⁷	7s ² 5f ⁷ 6d ¹	7s ² 5f ⁹ 6d ¹	7s ² 5f ¹⁰	7s ² 5f ¹¹	7s ² 5f ¹²	7s ² 5f ¹³	7s ² 5f ¹⁴	7s ² 5f ¹⁴ 6d ¹

Atomic number	→ 1	H	← Name
		1s ¹	← Outer electron configuration

This version of the periodic table shows the outer-shell electron configuration of each element. Note that down each group, the configuration is often similar.

When we come to the next element in the periodic table, the alkali metal potassium (atomic number 19), we might expect that we would begin to add electrons to the $3d$ subshell. However, all available chemical and physical evidence indicates that potassium is like lithium and sodium, and that the next electron is not added to the $3d$ level but is, instead, added to the $4s$ level ([\[link\]](#)). As discussed previously, the $3d$ orbital with no radial nodes is higher in energy because it is less penetrating and more shielded from the nucleus than the $4s$, which has three radial nodes. Thus, potassium has an electron configuration of $[\text{Ar}]4s^1$. Hence, potassium corresponds to Li and Na in its valence shell configuration. The next electron is added to complete the $4s$ subshell and calcium has an electron configuration of $[\text{Ar}]4s^2$. This gives calcium an outer-shell electron configuration corresponding to that of beryllium and magnesium.

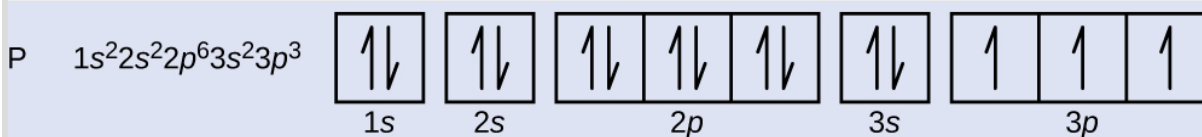
Beginning with the transition metal scandium (atomic number 21), additional electrons are added successively to the $3d$ subshell. This subshell is filled to its capacity with 10 electrons (remember that for $l = 2$ [d orbitals], there are $2l + 1 = 5$ values of m_l , meaning that there are five d orbitals that have a combined capacity of 10 electrons). The $4p$ subshell fills next. Note that for three series of elements, scandium (Sc) through copper (Cu), yttrium (Y) through silver (Ag), and lutetium (Lu) through gold (Au), a total of 10 d electrons are successively added to the $(n - 1)$ shell next to the n shell to bring that $(n - 1)$ shell from 8 to 18 electrons. For two series, lanthanum (La) through lutetium (Lu) and actinium (Ac) through lawrencium (Lr), 14 f electrons ($l = 3$, $2l + 1 = 7$ m_l values; thus, seven orbitals with a combined capacity of 14 electrons) are successively added to the $(n - 2)$ shell to bring that shell from 18 electrons to a total of 32 electrons.

Example:**Quantum Numbers and Electron Configurations**

What is the electron configuration and orbital diagram for a phosphorus atom? What are the four quantum numbers for the last electron added?

Solution

The atomic number of phosphorus is 15. Thus, a phosphorus atom contains 15 electrons. The order of filling of the energy levels is $1s$, $2s$, $2p$, $3s$, $3p$, $4s$, . . . The 15 electrons of the phosphorus atom will fill up to the $3p$ orbital, which will contain three electrons:



The last electron added is a $3p$ electron. Therefore, $n = 3$ and, for a p -type orbital, $l = 1$. The m_l value could be -1 , 0 , or $+1$. The three p orbitals are degenerate, so any of these m_l values is correct. For unpaired electrons, convention assigns the value of $+\frac{1}{2}$ for the spin quantum number; thus, $m_s = +\frac{1}{2}$.

Check Your Learning

Identify the atoms from the electron configurations given:

(a) $[\text{Ar}]4s^2 3d^5$

(b) $[\text{Kr}]5s^2 4d^{10} 5p^6$

Note:

Answer:

(a) Mn (b) Xe

The periodic table can be a powerful tool in predicting the electron configuration of an element. However, we do find exceptions to the order of filling of orbitals that are shown in [\[link\]](#) or [\[link\]](#). For instance, the electron configurations (shown in [\[link\]](#)) of the transition metals chromium (Cr; atomic number 24) and copper (Cu; atomic number 29), among others, are not those we would expect. In general, such exceptions involve subshells with very similar energy, and small effects can lead to changes in the order of filling.

In the case of Cr and Cu, we find that half-filled and completely filled subshells apparently represent conditions of preferred stability. This stability is such that an electron shifts from the 4s into the 3d orbital to gain the extra stability of a half-filled 3d subshell (in Cr) or a filled 3d subshell (in Cu). Other exceptions also occur. For example, niobium (Nb, atomic number 41) is predicted to have the electron configuration $[\text{Kr}]5s^24d^3$. Experimentally, we observe that its ground-state electron configuration is actually $[\text{Kr}]5s^14d^4$. We can rationalize this observation by saying that the electron–electron repulsions experienced by pairing the electrons in the 5s orbital are larger than the gap in energy between the 5s and 4d orbitals. There is no simple method to predict the exceptions for atoms where the magnitude of the repulsions between electrons is greater than the small differences in energy between subshells.

Electron Configurations and the Periodic Table

As described earlier, the periodic table arranges atoms based on increasing atomic number so that elements with the same chemical properties recur periodically. When their electron configurations are added to the table ([link](#)), we also see a periodic recurrence of similar electron configurations in the outer shells of these elements. Because they are in the outer shells of an atom, valence electrons play the most important role in chemical reactions. The outer electrons have the highest energy of the electrons in an atom and are more easily lost or shared than the core electrons. Valence electrons are also the determining factor in some physical properties of the elements.

Elements in any one group (or column) have the same number of valence electrons; the alkali metals lithium and sodium each have only one valence electron, the alkaline earth metals beryllium and magnesium each have two, and the halogens fluorine and chlorine each have seven valence electrons. The similarity in chemical properties among elements of the same group occurs because they have the same number of valence electrons. It is the loss, gain, or sharing of valence electrons that defines how elements react.

It is important to remember that the periodic table was developed on the basis of the chemical behavior of the elements, well before any idea of their

atomic structure was available. Now we can understand why the periodic table has the arrangement it has—the arrangement puts elements whose atoms have the same number of valence electrons in the same group. This arrangement is emphasized in [\[link\]](#), which shows in periodic-table form the electron configuration of the last subshell to be filled by the Aufbau principle. The colored sections of [\[link\]](#) show the three categories of elements classified by the orbitals being filled: main group, transition, and inner transition elements. These classifications determine which orbitals are counted in the **valence shell**, or highest energy level orbitals of an atom.

1. **Main group elements** (sometimes called **representative elements**) are those in which the last electron added enters an s or a p orbital in the outermost shell, shown in blue and red in [\[link\]](#). This category includes all the nonmetallic elements, as well as many metals and the metalloids. The valence electrons for main group elements are those with the highest n level. For example, gallium (Ga, atomic number 31) has the electron configuration $[\text{Ar}]\underline{4s^2}3d^{10}\underline{4p^1}$, which contains three valence electrons (underlined). The completely filled d orbitals count as core, not valence, electrons.
2. **Transition elements or transition metals.** These are metallic elements in which the last electron added enters a d orbital. The valence electrons (those added after the last noble gas configuration) in these elements include the ns and $(n - 1) d$ electrons. The official IUPAC definition of transition elements specifies those with partially filled d orbitals. Thus, the elements with completely filled orbitals (Zn, Cd, Hg, as well as Cu, Ag, and Au in [\[link\]](#)) are not technically transition elements. However, the term is frequently used to refer to the entire d block (colored yellow in [\[link\]](#)), and we will adopt this usage in this textbook.
3. **Inner transition elements** are metallic elements in which the last electron added occupies an f orbital. They are shown in green in [\[link\]](#). The valence shells of the inner transition elements consist of the $(n - 2)f$, the $(n - 1)d$, and the ns subshells. There are two inner transition series:
 - a. The lanthanide series: lanthanide (La) through lutetium (Lu)
 - b. The actinide series: actinide (Ac) through lawrencium (Lr)

Lanthanum and actinium, because of their similarities to the other members of the series, are included and used to name the series, even though they are transition metals with no f electrons.

Electron Configurations of Ions

Ions are formed when atoms gain or lose electrons. A cation (positively charged ion) forms when one or more electrons are removed from a parent atom. For main group elements, the electrons that were added last are the first electrons removed. For transition metals and inner transition metals, however, electrons in the s orbital are easier to remove than the d or f electrons, and so the highest ns electrons are lost, and then the $(n - 1)d$ or $(n - 2)f$ electrons are removed. An anion (negatively charged ion) forms when one or more electrons are added to a parent atom. The added electrons fill in the order predicted by the Aufbau principle.

Example:

Predicting Electron Configurations of Ions

What is the electron configuration of:

- (a) Na^+
- (b) P^{3-}
- (c) Al^{2+}
- (d) Fe^{2+}
- (e) Sm^{3+}

Solution

First, write out the electron configuration for each parent atom. We have chosen to show the full, unabbreviated configurations to provide more practice for students who want it, but listing the core-abbreviated electron configurations is also acceptable.

Next, determine whether an electron is gained or lost. Remember electrons are negatively charged, so ions with a positive charge have *lost* an electron. For main group elements, the last orbital gains or loses the electron. For transition metals, the last s orbital loses an electron before the d orbitals.

(a) Na: $1s^2 2s^2 2p^6 3s^1$. Sodium cation loses one electron, so Na^+ :
 $1s^2 2s^2 2p^6 3s^1 = \text{Na}^+ : 1s^2 2s^2 2p^6$.

(b) P: $1s^2 2s^2 2p^6 3s^2 3p^3$. Phosphorus trianion gains three electrons, so P^{3-} : $1s^2 2s^2 2p^6 3s^2 3p^6$.

(c) Al: $1s^2 2s^2 2p^6 3s^2 3p^1$. Aluminum dication loses two electrons Al^{2+} : $1s^2 2s^2 2p^6 3s^2 3p^1 =$
 Al^{2+} : $1s^2 2s^2 2p^6 3s^1$.

(d) Fe: $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^6$. Iron(II) loses two electrons and, since it is a transition metal, they are removed from the 4s orbital Fe^{2+} :

$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^6 = 1s^2 2s^2 2p^6 3s^2 3p^6 3d^6$.

(e). Sm: $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10} 4p^6 5s^2 4d^{10} 5p^6 6s^2 4f^6$. Samarium trication loses three electrons. The first two will be lost from the 6s orbital, and the final one is removed from the 4f orbital. Sm^{3+} :

$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10} 4p^6 5s^2 4d^{10} 5p^6 6s^2 4f^6 =$

$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10} 4p^6 5s^2 4d^{10} 5p^6 4f^5$.

Check Your Learning

Which ion with a +2 charge has the electron configuration

$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^5$? Which ion with a +3 charge has this configuration?

Note:

Answer:

Tc^{2+} , Ru^{3+}

Key Concepts and Summary

The relative energy of the subshells determine the order in which atomic orbitals are filled (1s, 2s, 2p, 3s, 3p, 4s, 3d, 4p, and so on). Electron configurations and orbital diagrams can be determined by applying the Pauli exclusion principle (no two electrons can have the same set of four quantum numbers) and Hund's rule (whenever possible, electrons retain unpaired spins in degenerate orbitals).

Electrons in the outermost orbitals, called valence electrons, are responsible for most of the chemical behavior of elements. In the periodic table, elements with analogous valence electron configurations usually occur within the same group. There are some exceptions to the predicted filling order, particularly when half-filled or completely filled orbitals can be formed. The periodic table can be divided into three categories based on the orbital in which the last electron to be added is placed: main group elements (*s* and *p* orbitals), transition elements (*d* orbitals), and inner transition elements (*f* orbitals).

Chemistry End of Chapter Exercises

Exercise:

Problem:

Read the labels of several commercial products and identify monatomic ions of at least four transition elements contained in the products. Write the complete electron configurations of these cations.

Exercise:

Problem:

Read the labels of several commercial products and identify monatomic ions of at least six main group elements contained in the products. Write the complete electron configurations of these cations and anions.

Solution:

For example, Na^+ : $1s^2 2s^2 2p^6$; Ca^{2+} : $1s^2 2s^2 2p^6$; Sn^{2+} : $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 5s^2$; F^- : $1s^2 2s^2 2p^6$; O^{2-} : $1s^2 2s^2 2p^6$; Cl^- : $1s^2 2s^2 2p^6 3s^2 3p^6$.

Exercise:

Problem:

Using complete subshell notation (not abbreviations, $1s^22s^22p^6$, and so forth), predict the electron configuration of each of the following atoms:

- (a) C
- (b) P
- (c) V
- (d) Sb
- (e) Sm

Exercise:**Problem:**

Using complete subshell notation ($1s^22s^22p^6$, and so forth), predict the electron configuration of each of the following atoms:

- (a) N
- (b) Si
- (c) Fe
- (d) Te
- (e) Tb

Solution:

(a) $1s^22s^22p^3$; (b) $1s^22s^22p^63s^23p^2$; (c) $1s^22s^22p^63s^23p^64s^23d^6$; (d) $1s^22s^22p^63s^23p^64s^23d^{10}4p^65s^24d^{10}5p^4$; (e) $1s^22s^22p^63s^23p^64s^23d^{10}4p^65s^24d^{10}5p^66s^24f^9$

Exercise:

Problem:

Is $1s^22s^22p^6$ the symbol for a macroscopic property or a microscopic property of an element? Explain your answer.

Exercise:**Problem:**

What additional information do we need to answer the question “Which ion has the electron configuration $1s^22s^22p^63s^23p^6$ ”?

Solution:

The charge on the ion.

Exercise:**Problem:**

Draw the orbital diagram for the valence shell of each of the following atoms:

(a) C

(b) P

(c) V

(d) Sb

(e) Ru

Exercise:**Problem:**

Use an orbital diagram to describe the electron configuration of the valence shell of each of the following atoms:

(a) N

(b) Si

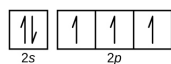
(c) Fe

(d) Te

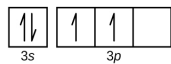
(e) Mo

Solution:

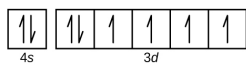
(a)



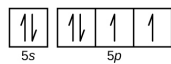
(b)



(c)



(d)



(e)



Exercise:

Problem:

Using complete subshell notation ($1s^2 2s^2 2p^6$, and so forth), predict the electron configurations of the following ions.

(a) N^{3-}

(b) Ca^{2+}

(c) S^-

(d) Cs^{2+}

(e) Cr^{2+}

(f) Gd^{3+}

Exercise:**Problem:**

Which atom has the electron configuration $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10} 4p^6 5s^2 4d^2$?

Solution:

Zr

Exercise:**Problem:**

Which atom has the electron configuration $1s^2 2s^2 2p^6 3s^2 3p^6 3d^7 4s^2$?

Exercise:**Problem:**

Which ion with a +1 charge has the electron configuration $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6$? Which ion with a -2 charge has this configuration?

Solution:



Exercise:

Problem:

Which of the following atoms contains only three valence electrons:
Li, B, N, F, Ne?

Exercise:

Problem: Which of the following has two unpaired electrons?

- (a) Mg
 - (b) Si
 - (c) S
 - (d) Both Mg and S
 - (e) Both Si and S.
-

Solution:

Although both (b) and (c) are correct, (e) encompasses both and is the best answer.

Exercise:

Problem:

Which atom would be expected to have a half-filled $6p$ subshell?

Exercise:

Problem:

Which atom would be expected to have a half-filled $4s$ subshell?

Solution:

K

Exercise:**Problem:**

In one area of Australia, the cattle did not thrive despite the presence of suitable forage. An investigation showed the cause to be the absence of sufficient cobalt in the soil. Cobalt forms cations in two oxidation states, Co^{2+} and Co^{3+} . Write the electron structure of the two cations.

Exercise:**Problem:**

Thallium was used as a poison in the Agatha Christie mystery story "The Pale Horse." Thallium has two possible cationic forms, +1 and +3. The +1 compounds are the more stable. Write the electron structure of the +1 cation of thallium.

Solution:

$1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 5s^2 5p^6 6s^2 4f^{14} 5d^{10}$

Exercise:**Problem:**

Write the electron configurations for the following atoms or ions:

(a) B^{3+}

(b) O^-

(c) Cl^{3+}

(d) Ca^{2+}

(e) Ti

Exercise:**Problem:**

Cobalt-60 and iodine-131 are radioactive isotopes commonly used in nuclear medicine. How many protons, neutrons, and electrons are in atoms of these isotopes? Write the complete electron configuration for each isotope.

Solution:

Co has 27 protons, 27 electrons, and 33 neutrons:
 $1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^7$. I has 53 protons, 53 electrons, and 78 neutrons: $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10} 5s^2 5p^5$.

Exercise:**Problem:**

Write a set of quantum numbers for each of the electrons with an n of 3 in a Sc atom.

Glossary**Aufbau principle**

procedure in which the electron configuration of the elements is determined by “building” them in order of atomic numbers, adding one proton to the nucleus and one electron to the proper subshell at a time

core electron

electron in an atom that occupies the orbitals of the inner shells

electron configuration

listing that identifies the electron occupancy of an atom’s shells and subshells

Hund’s rule

every orbital in a subshell is singly occupied with one electron before any one orbital is doubly occupied, and all electrons in singly occupied

orbitals have the same spin

orbital diagram

pictorial representation of the electron configuration showing each orbital as a box and each electron as an arrow

valence electrons

electrons in the outermost or valence shell (highest value of n) of a ground-state atom

valence shell

outermost shell of electrons in a ground-state atom

Periodic Variations in Element Properties

By the end of this section, you will be able to:

- Describe and explain the observed trends in atomic size, ionization energy, and electron affinity of the elements

The elements in groups (vertical columns) of the periodic table exhibit similar chemical behavior. This similarity occurs because the members of a group have the same number and distribution of electrons in their valence shells. However, there are also other patterns in chemical properties on the periodic table. For example, as we move down a group, the metallic character of the atoms increases. Oxygen, at the top of group 16 (6A), is a colorless gas; in the middle of the group, selenium is a semiconducting solid; and, toward the bottom, polonium is a silver-grey solid that conducts electricity.

As we go across a period from left to right, we add a proton to the nucleus and an electron to the valence shell with each successive element. As we go down the elements in a group, the number of electrons in the valence shell remains constant, but the principal quantum number increases by one each time. An understanding of the electronic structure of the elements allows us to examine some of the properties that govern their chemical behavior. These properties vary periodically as the electronic structure of the elements changes. They are (1) size (radius) of atoms and ions, (2) ionization energies, and (3) electron affinities.

Note:

Explore [visualizations](#) of the periodic trends discussed in this section (and many more trends). With just a few clicks, you can create three-dimensional versions of the periodic table showing atomic size or graphs of ionization energies from all measured elements.

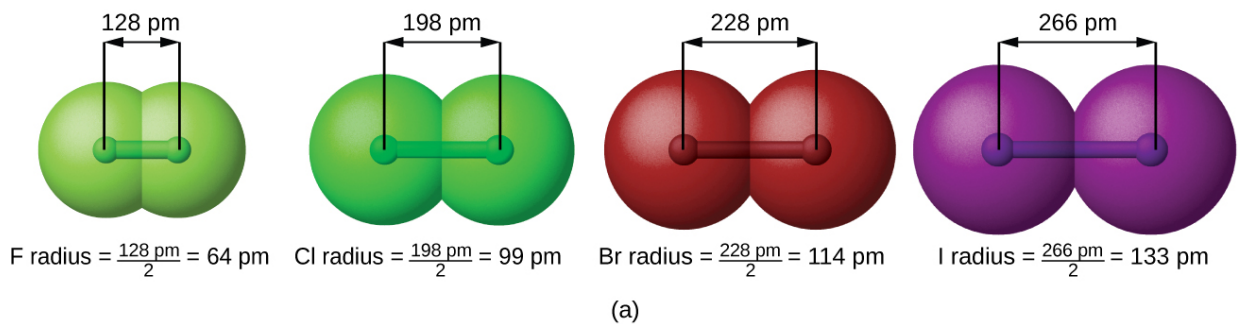
Variation in Covalent Radius

The quantum mechanical picture makes it difficult to establish a definite size of an atom. However, there are several practical ways to define the radius of atoms and, thus, to determine their relative sizes that give roughly similar values. We will use the **covalent radius** ([link](#)), which is defined as one-half the distance between the nuclei of two identical atoms when they are joined by a covalent bond (this measurement is possible because atoms within molecules still retain much of their atomic identity). We know that as we scan down a group, the principal quantum number, n , increases by one for each element. Thus, the electrons are being added to a region of space that is increasingly distant from the nucleus. Consequently, the size of the atom (and its covalent radius) must increase as we increase the distance of the outermost electrons from the nucleus. This trend is illustrated for the covalent radii of the halogens in [link](#) and [link](#). The trends for the entire periodic table can be seen in [link](#).

Covalent Radii of the Halogen Group Elements

Atomic Radii of the Halogens (pm)	Nuclear charge
-----------------------------------	----------------

Atom	Covalent radius (pm)	Nuclear charge
F	64	+9
Cl	99	+17
Br	114	+35
I	133	+53
At	148	+85



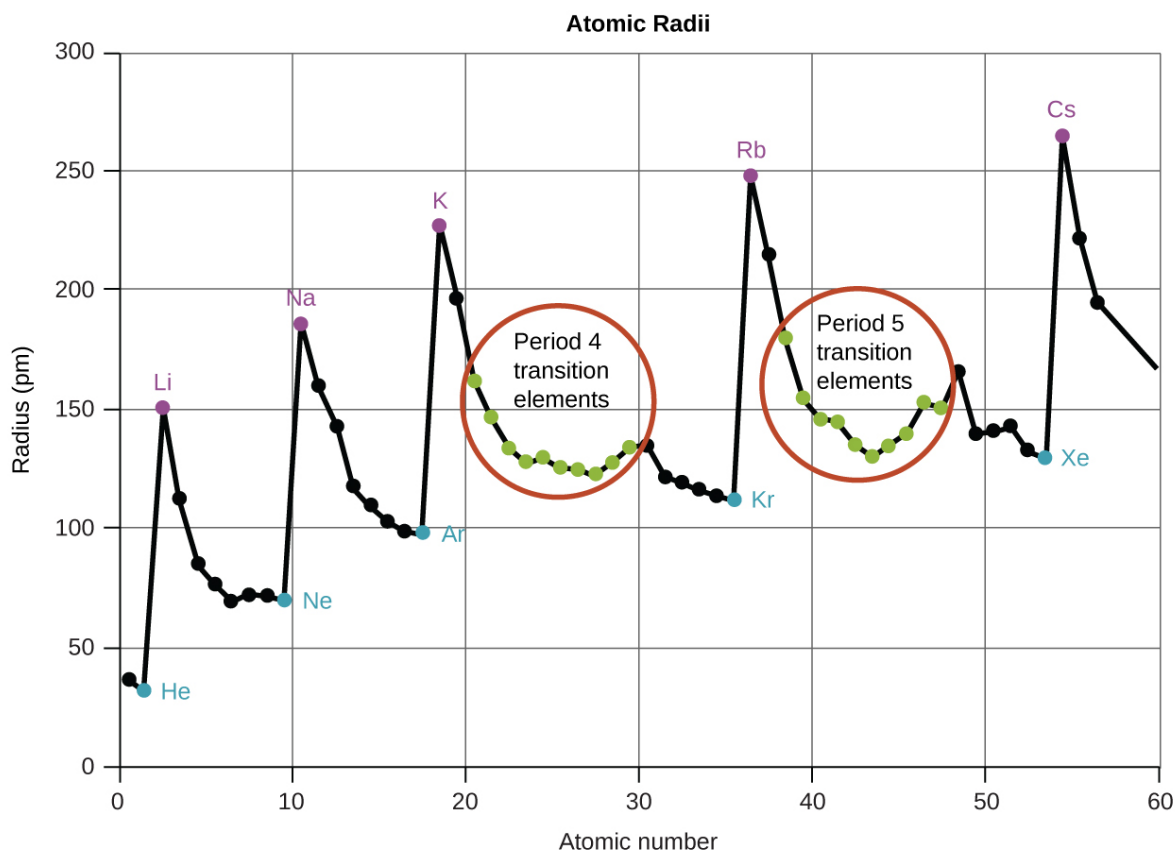
Periodic Table of the Elements

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	H																	He
2	Li	Be											B	C	N	O	F	Ne
3	Na	Mg											Al	Si	P	S	Cl	Ar
4	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
5	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
6	Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
7	Fr	Ra	Ac	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Cn	Nh	Fl	Mc	Lv	Ts	Og

(b)

(a) The radius of an atom is defined as one-half the distance between the nuclei in a molecule

consisting of two identical atoms joined by a covalent bond. The atomic radius for the halogens increases down the group as n increases. (b) Covalent radii of the elements are shown to scale. The general trend is that radii increase down a group and decrease across a period.



Within each period, the trend in atomic radius decreases as Z increases; for example, from K to Kr. Within each group (e.g., the alkali metals shown in purple), the trend is that atomic radius increases as Z increases.

As shown in [\[link\]](#), as we move across a period from left to right, we generally find that each element has a smaller covalent radius than the element preceding it. This might seem counterintuitive because it implies that atoms with more electrons have a smaller atomic radius. This can be explained with the concept of **effective nuclear charge, Z_{eff}** . This is the pull exerted on a specific electron by the nucleus, taking into account any electron–electron repulsions. For hydrogen, there is only one electron and so the nuclear charge (Z) and the effective nuclear charge (Z_{eff}) are equal. For all other atoms, the inner electrons partially *shield* the outer electrons from the pull of the nucleus, and thus:

Equation:

$$Z_{\text{eff}} = Z - \text{shielding}$$

Shielding is determined by the probability of another electron being between the electron of interest and the nucleus, as well as by the electron–electron repulsions the electron of interest encounters. Core electrons are adept at shielding, while electrons in the same valence shell do not block the nuclear attraction experienced by each other as efficiently. Thus, each time we move from one element to the next across a period, Z increases by one, but the shielding increases only slightly. Thus, Z_{eff} increases as we move from left to right across a period. The stronger pull (higher effective nuclear charge) experienced by electrons on the right side of the periodic table draws them closer to the nucleus, making the covalent radii smaller.

Thus, as we would expect, the outermost or valence electrons are easiest to remove because they have the highest energies, are shielded more, and are farthest from the nucleus. As a general rule, when the representative elements form cations, they do so by the loss of the ns or np electrons that were added last in the Aufbau process. The transition elements, on the other hand, lose the ns electrons before they begin to lose the $(n - 1)d$ electrons, even though the ns electrons are added first, according to the Aufbau principle.

Example:

Sorting Atomic Radii

Predict the order of increasing covalent radius for Ge, Fl, Br, Kr.

Solution

Radius increases as we move down a group, so $\text{Ge} < \text{Fl}$ (Note: Fl is the symbol for flerovium, element 114, NOT fluorine). Radius decreases as we move across a period, so $\text{Kr} < \text{Br} < \text{Ge}$. Putting the trends together, we obtain $\text{Kr} < \text{Br} < \text{Ge} < \text{Fl}$.

Check Your Learning

Give an example of an atom whose size is smaller than fluorine.

Note:

Answer:

Ne or He

Variation in Ionic Radii

Ionic radius is the measure used to describe the size of an ion. A cation always has fewer electrons and the same number of protons as the parent atom; it is smaller than the atom from which it is derived ([\[link\]](#)). For example, the covalent radius of an aluminum atom ($1s^2 2s^2 2p^6 3s^2 3p^1$) is 118 pm, whereas the ionic radius of an Al^{3+} ($1s^2 2s^2 2p^6$) is 68 pm. As electrons are removed from the outer valence shell, the remaining core electrons occupying smaller shells experience a greater effective nuclear charge Z_{eff} (as discussed) and are drawn even closer to the nucleus.



The radius for a cation is smaller than the parent atom (Al), due to the lost electrons; the radius for an anion is larger than the parent (S), due to the gained electrons.

Cations with larger charges are smaller than cations with smaller charges (e.g., V²⁺ has an ionic radius of 79 pm, while that of V³⁺ is 64 pm). Proceeding down the groups of the periodic table, we find that cations of successive elements with the same charge generally have larger radii, corresponding to an increase in the principal quantum number, n .

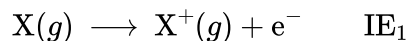
An anion (negative ion) is formed by the addition of one or more electrons to the valence shell of an atom. This results in a greater repulsion among the electrons and a decrease in Z_{eff} per electron. Both effects (the increased number of electrons and the decreased Z_{eff}) cause the radius of an anion to be larger than that of the parent atom ([link](#)). For example, a sulfur atom ([Ne]3s²3p⁴) has a covalent radius of 104 pm, whereas the ionic radius of the sulfide anion ([Ne]3s²3p⁶) is 170 pm. For consecutive elements proceeding down any group, anions have larger principal quantum numbers and, thus, larger radii.

Atoms and ions that have the same electron configuration are said to be **isoelectronic**. Examples of isoelectronic species are N³⁻, O²⁻, F⁻, Ne, Na⁺, Mg²⁺, and Al³⁺ (1s²2s²2p⁶). Another isoelectronic series is P³⁻, S²⁻, Cl⁻, Ar, K⁺, Ca²⁺, and Sc³⁺ ([Ne]3s²3p⁶). For atoms or ions that are isoelectronic, the number of protons determines the size. The greater the nuclear charge, the smaller the radius in a series of isoelectronic ions and atoms.

Variation in Ionization Energies

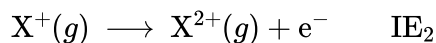
The amount of energy required to remove the most loosely bound electron from a gaseous atom in its ground state is called its first **ionization energy** (IE₁). The first ionization energy for an element, X, is the energy required to form a cation with +1 charge:

Equation:



The energy required to remove the second most loosely bound electron is called the second ionization energy (IE₂).

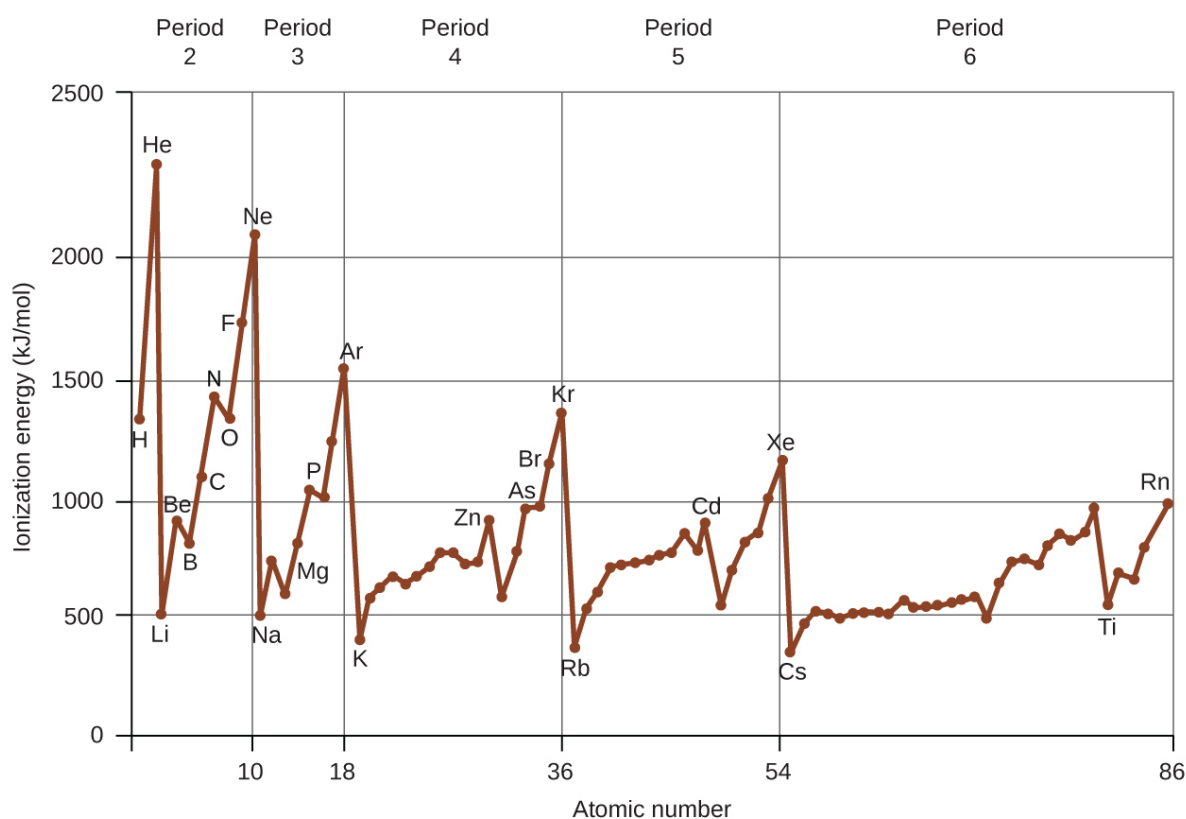
Equation:



The energy required to remove the third electron is the third ionization energy, and so on. Energy is always required to remove electrons from atoms or ions, so ionization processes are endothermic and IE values are always positive. For larger atoms, the most loosely bound electron is located farther from the nucleus and so is easier to remove. Thus, as size (atomic radius) increases, the ionization

energy should decrease. Relating this logic to what we have just learned about radii, we would expect first ionization energies to decrease down a group and to increase across a period.

[\[link\]](#) graphs the relationship between the first ionization energy and the atomic number of several elements. The values of first ionization energy for the elements are given in [\[link\]](#). Within a period, the IE_1 generally increases with increasing Z . Down a group, the IE_1 value generally decreases with increasing Z . There are some systematic deviations from this trend, however. Note that the ionization energy of boron (atomic number 5) is less than that of beryllium (atomic number 4) even though the nuclear charge of boron is greater by one proton. This can be explained because the energy of the subshells increases as l increases, due to penetration and shielding (as discussed previously in this chapter). Within any one shell, the s electrons are lower in energy than the p electrons. This means that an s electron is harder to remove from an atom than a p electron in the same shell. The electron removed during the ionization of beryllium ($[\text{He}]2s^2$) is an s electron, whereas the electron removed during the ionization of boron ($[\text{He}]2s^22p^1$) is a p electron; this results in a lower first ionization energy for boron, even though its nuclear charge is greater by one proton. Thus, we see a small deviation from the predicted trend occurring each time a new subshell begins.

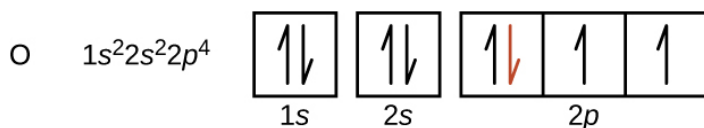


The first ionization energy of the elements in the first five periods are plotted against their atomic number.

		First Ionization Energies of Some Elements (kJ/mol)																	
Period	Group																		
	1											13	14	15	16	17	18		
1	H 1310																	He 2370	
2	Li 520	Be 900											B 800	C 1090	N 1400	O 1310	F 1680	Ne 2080	
3	Na 490	Mg 730											Al 580	Si 780	P 1060	S 1000	Cl 1250	Ar 1520	
4	K 420	Ca 590	Sc 630	Ti 660	V 650	Cr 660	Mn 710	Fe 760	Co 760	Ni 730	Cu 740	Zn 910	Ga 580	Ge 780	As 960	Se 950	Br 1140	Kr 1350	
5	Rb 400	Sr 550	Y 620	Zr 660	Nb 670	Mo 680	Tc 700	Ru 710	Rh 720	Pd 800	Ag 730	Cd 870	In 560	Sn 700	Sb 830	Te 870	I 1010	Xe 1170	
6	Cs 380	Ba 500	La 540	Hf 700	Ta 760	W 770	Re 760	Os 840	Ir 890	Pt 870	Au 890	Hg 1000	Tl 590	Pb 710	Bi 800	Po 810	At ...	Rn 1030	
7	Fr ...	Ra 510																	

This version of the periodic table shows the first ionization energy (IE_1), in kJ/mol, of selected elements.

Another deviation occurs as orbitals become more than one-half filled. The first ionization energy for oxygen is slightly less than that for nitrogen, despite the trend in increasing IE_1 values across a period. Looking at the orbital diagram of oxygen, we can see that removing one electron will eliminate the electron–electron repulsion caused by pairing the electrons in the $2p$ orbital and will result in a half-filled orbital (which is energetically favorable). Analogous changes occur in succeeding periods (note the dip for sulfur after phosphorus in [\[link\]](#)).



Removing an electron from a cation is more difficult than removing an electron from a neutral atom because of the greater electrostatic attraction to the cation. Likewise, removing an electron from a cation with a higher positive charge is more difficult than removing an electron from an ion with a lower charge. Thus, successive ionization energies for one element always increase. As seen in [\[link\]](#), there is a large increase in the ionization energies for each element. This jump corresponds to removal of the core electrons, which are harder to remove than the valence electrons. For example, Sc and Ga both have three valence electrons, so the rapid increase in ionization energy occurs after the third ionization.

Successive Ionization Energies for Selected Elements (kJ/mol)

Successive Ionization Energies for Selected Elements (kJ/mol)	IE ₁	IE ₂	IE ₃	IE ₄	IE ₅	IE ₆	IE ₇
Element	IE ₁	IE ₂	IE ₃	IE ₄	IE ₅	IE ₆	IE ₇
K	418.8	3051.8	4419.6	5876.9	7975.5	9590.6	11343
Ca	589.8	1145.4	4912.4	6490.6	8153.0	10495.7	12272.9
Sc	633.1	1235.0	2388.7	7090.6	8842.9	10679.0	13315.0
Ga	578.8	1979.4	2964.6	6180	8298.7	10873.9	13594.8
Ge	762.2	1537.5	3302.1	4410.6	9021.4	Not available	Not available
As	944.5	1793.6	2735.5	4836.8	6042.9	12311.5	Not available

Example:

Ranking Ionization Energies

Predict the order of increasing energy for the following processes: IE₁ for Al, IE₁ for Tl, IE₂ for Na, IE₃ for Al.

Solution

Removing the $6p^1$ electron from Tl is easier than removing the $3p^1$ electron from Al because the higher n orbital is farther from the nucleus, so $IE_1(\text{Tl}) < IE_1(\text{Al})$. Ionizing the third electron from Al ($\text{Al}^{2+} \rightarrow \text{Al}^{3+} + e^-$) requires more energy because the cation Al^{2+} exerts a stronger pull on the electron than the neutral Al atom, so $IE_1(\text{Al}) < IE_3(\text{Al})$. The second ionization energy for sodium removes a core electron, which is a much higher energy process than removing valence electrons. Putting this all together, we obtain: $IE_1(\text{Tl}) < IE_1(\text{Al}) < IE_3(\text{Al}) < IE_2(\text{Na})$.

Check Your Learning

Which has the lowest value for IE₁: O, Po, Pb, or Ba?

Note:

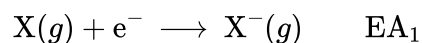
Answer:

Ba

Variation in Electron Affinities

The **electron affinity** (EA) is the energy change for the process of adding an electron to a gaseous atom to form an anion (negative ion).

Equation:



This process can be either endothermic or exothermic, depending on the element. The EA of some of the elements is given in [\[link\]](#). You can see that many of these elements have negative values of EA, which means that energy is released when the gaseous atom accepts an electron. However, for some elements, energy is required for the atom to become negatively charged and the value of their EA is positive. Just as with ionization energy, subsequent EA values are associated with forming ions with more charge. The second EA is the energy associated with adding an electron to an anion to form a -2 ion, and so on.

As we might predict, it becomes easier to add an electron across a series of atoms as the effective nuclear charge of the atoms increases. We find, as we go from left to right across a period, EAs tend to become more negative. The exceptions found among the elements of group 2 (2A), group 15 (5A), and group 18 (8A) can be understood based on the electronic structure of these groups. The noble gases, group 18 (8A), have a completely filled shell and the incoming electron must be added to a higher n level, which is more difficult to do. Group 2 (2A) has a filled ns subshell, and so the next electron added goes into the higher energy np , so, again, the observed EA value is not as the trend would predict. Finally, group 15 (5A) has a half-filled np subshell and the next electron must be paired with an existing np electron. In all of these cases, the initial relative stability of the electron configuration disrupts the trend in EA.

We also might expect the atom at the top of each group to have the most negative EA; their first ionization potentials suggest that these atoms have the largest effective nuclear charges. However, as we move down a group, we see that the *second* element in the group most often has the most negative EA. This can be attributed to the small size of the $n = 2$ shell and the resulting large electron–electron repulsions. For example, chlorine, with an EA value of -348 kJ/mol, has the highest value of any element in the periodic table. The EA of fluorine is -322 kJ/mol. When we add an electron to a fluorine atom to form a fluoride anion (F^-), we add an electron to the $n = 2$ shell. The electron is attracted to the nucleus, but there is also significant repulsion from the other electrons already present in this small valence shell. The chlorine atom has the same electron configuration in the valence shell, but because the entering electron is going into the $n = 3$ shell, it occupies a considerably larger region of space and the electron–electron repulsions are reduced. The entering electron does not experience as much repulsion and the chlorine atom accepts an additional electron more readily, resulting in a more negative EA.

Period		Electron Affinity Values for Selected Elements (kJ/mol)																Group	
		1	2											13	14	15	16	17	18
1		H -72																	He +20*
2		Li -60	Be +240*											B -23	C -123	N 0	O -141	F -322	Ne -30
3		Na -53	Mg +230*											Al -44	Si -120	P -74	S -20	Cl -348	Ar +35*
4		K -48	Ca +150*	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga -40*	Ge -115	As -7	Se -195	Br -324	Kr +40*
5		Rb -46	Sr +160*	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In -40*	Sn -121	Sb -101	Te -190	I -295	Xe +40*
6		Cs -45	Ba +50*	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl -50	Pb -101	Bi -101	Po -170	At -270*	Rn +40*
7		Fr	Ra																

* Calculated value

This version of the periodic table displays the electron affinity values (in kJ/mol) for selected elements.

The properties discussed in this section (size of atoms and ions, effective nuclear charge, ionization energies, and electron affinities) are central to understanding chemical reactivity. For example, because fluorine has an energetically favorable EA and a large energy barrier to ionization (IE), it is much easier to form fluorine anions than cations. Metallic properties including conductivity and malleability (the ability to be formed into sheets) depend on having electrons that can be removed easily. Thus, metallic character increases as we move down a group and decreases across a period in the same trend observed for atomic size because it is easier to remove an electron that is farther away from the nucleus.

Key Concepts and Summary

Electron configurations allow us to understand many periodic trends. Covalent radius increases as we move down a group because the n level (orbital size) increases. Covalent radius mostly decreases as we move left to right across a period because the effective nuclear charge experienced by the electrons increases, and the electrons are pulled in tighter to the nucleus. Anionic radii are larger than the parent atom, while cationic radii are smaller, because the number of valence electrons has changed while the nuclear charge has remained constant. Ionization energy (the energy associated with forming a cation) decreases down a group and mostly increases across a period because it is easier to remove an electron from a larger, higher energy orbital. Electron affinity (the energy associated with forming an anion) is more favorable (exothermic) when electrons are placed into lower energy orbitals, closer to the nucleus. Therefore, electron affinity becomes increasingly negative as we move left to right across the periodic table and decreases as we move down a group. For both IE and electron affinity data, there are exceptions to the trends when dealing with completely filled or half-filled subshells.

Chemistry End of Chapter Exercises

Exercise:

Problem:

Based on their positions in the periodic table, predict which has the smallest atomic radius: Mg, Sr, Si, Cl, I.

Solution:

Cl

Exercise:

Problem:

Based on their positions in the periodic table, predict which has the largest atomic radius: Li, Rb, N, F, I.

Exercise:

Problem:

Based on their positions in the periodic table, predict which has the largest first ionization energy: Mg, Ba, B, O, Te.

Solution:

O

Exercise:

Problem:

Based on their positions in the periodic table, predict which has the smallest first ionization energy: Li, Cs, N, F, I.

Exercise:

Problem:

Based on their positions in the periodic table, rank the following atoms in order of increasing first ionization energy: F, Li, N, Rb

Solution:

Rb < Li < N < F

Exercise:

Problem:

Based on their positions in the periodic table, rank the following atoms in order of increasing first ionization energy: Mg, O, S, Si

Exercise:

Problem:

Atoms of which group in the periodic table have a valence shell electron configuration of ns^2np^3 ?

Solution:

15 (5A)

Exercise:**Problem:**

Atoms of which group in the periodic table have a valence shell electron configuration of ns^2 ?

Exercise:**Problem:**

Based on their positions in the periodic table, list the following atoms in order of increasing radius: Mg, Ca, Rb, Cs.

Solution:

$\text{Mg} < \text{Ca} < \text{Rb} < \text{Cs}$

Exercise:**Problem:**

Based on their positions in the periodic table, list the following atoms in order of increasing radius: Sr, Ca, Si, Cl.

Exercise:**Problem:**

Based on their positions in the periodic table, list the following ions in order of increasing radius: K^+ , Ca^{2+} , Al^{3+} , Si^{4+} .

Solution:

$\text{Si}^{4+} < \text{Al}^{3+} < \text{Ca}^{2+} < \text{K}^+$

Exercise:

Problem: List the following ions in order of increasing radius: Li^+ , Mg^{2+} , Br^- , Te^{2-} .

Exercise:

Problem: Which atom and/or ion is (are) isoelectronic with Br^+ : Se^{2+} , Se, As^- , Kr, Ga^{3+} , Cl^- ?

Solution:

Se, As^-

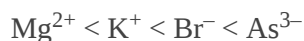
Exercise:

Problem:

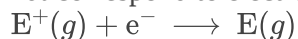
Which of the following atoms and ions is (are) isoelectronic with S^{2+} : Si^{4+} , Cl^{3+} , Ar, As^{3+} , Si, Al^{3+} ?

Exercise:**Problem:**

Compare both the numbers of protons and electrons present in each to rank the following ions in order of increasing radius: As^{3-} , Br^{-} , K^{+} , Mg^{2+} .

Solution:**Exercise:****Problem:**

Of the five elements Al, Cl, I, Na, Rb, which has the most exothermic reaction? (E represents an atom.) What name is given to the energy for the reaction? Hint: Note the process depicted does *not* correspond to electron affinity.)

**Exercise:****Problem:**

Of the five elements Sn, Si, Sb, O, Te, which has the most endothermic reaction? (E represents an atom.) What name is given to the energy for the reaction?

**Solution:**

O, IE_1

Exercise:**Problem:**

The ionic radii of the ions S^{2-} , Cl^{-} , and K^{+} are 184, 181, 138 pm respectively. Explain why these ions have different sizes even though they contain the same number of electrons.

Exercise:**Problem:**

Which main group atom would be expected to have the lowest second ionization energy?

Solution:

Ra

Exercise:

Problem: Explain why Al is a member of group 13 rather than group 3?

Glossary

covalent radius

one-half the distance between the nuclei of two identical atoms when they are joined by a covalent bond

effective nuclear charge

charge that leads to the Coulomb force exerted by the nucleus on an electron, calculated as the nuclear charge minus shielding

electron affinity

energy change associated with addition of an electron to a gaseous atom or ion

ionization energy

energy required to remove an electron from a gaseous atom or ion

isoelectronic

group of ions or atoms that have identical electron configurations

Chemical Reaction Rates

By the end of this section, you will be able to:

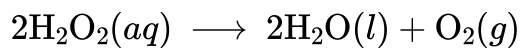
- Define chemical reaction rate
- Derive rate expressions from the balanced equation for a given chemical reaction
- Calculate reaction rates from experimental data

A *rate* is a measure of how some property varies with time. Speed is a familiar rate that expresses the distance traveled by an object in a given amount of time. Wage is a rate that represents the amount of money earned by a person working for a given amount of time. Likewise, the rate of a chemical reaction is a measure of how much reactant is consumed, or how much product is produced, by the reaction in a given amount of time.

The **rate of reaction** is the change in the amount of a reactant or product per unit time. Reaction rates are therefore determined by measuring the time dependence of some property that can be related to reactant or product amounts. Rates of reactions that consume or produce gaseous substances, for example, are conveniently determined by measuring changes in volume or pressure. For reactions involving one or more colored substances, rates may be monitored via measurements of light absorption. For reactions involving aqueous electrolytes, rates may be measured via changes in a solution's conductivity.

For reactants and products in solution, their relative amounts (concentrations) are conveniently used for purposes of expressing reaction rates. For example, the concentration of hydrogen peroxide, H_2O_2 , in an aqueous solution changes slowly over time as it decomposes according to the equation:

Equation:



The rate at which the hydrogen peroxide decomposes can be expressed in terms of the rate of change of its concentration, as shown here:

Equation:

$$\begin{aligned}\text{rate of decomposition of } \text{H}_2\text{O}_2 &= - \frac{\text{change in concentration of reactant}}{\text{time interval}} \\ &= - \frac{[\text{H}_2\text{O}_2]_{t_2} - [\text{H}_2\text{O}_2]_{t_1}}{t_2 - t_1} \\ &= - \frac{\Delta[\text{H}_2\text{O}_2]}{\Delta t}\end{aligned}$$

This mathematical representation of the change in species concentration over time is the **rate expression** for the reaction. The brackets indicate molar concentrations, and the

symbol delta (Δ) indicates “change in.” Thus, $[\text{H}_2\text{O}_2]_{t_1}$ represents the molar concentration of hydrogen peroxide at some time t_1 ; likewise, $[\text{H}_2\text{O}_2]_{t_2}$ represents the molar concentration of hydrogen peroxide at a later time t_2 ; and $\Delta[\text{H}_2\text{O}_2]$ represents the change in molar concentration of hydrogen peroxide during the time interval Δt (that is, $t_2 - t_1$). Since the reactant concentration decreases as the reaction proceeds, $\Delta[\text{H}_2\text{O}_2]$ is a negative quantity. Reaction rates are, by convention, positive quantities, and so this negative change in concentration is multiplied by -1 . [\[link\]](#) provides an example of data collected during the decomposition of H_2O_2 .

Time (h)	$[\text{H}_2\text{O}_2]$ (mol L ⁻¹)	$\Delta[\text{H}_2\text{O}_2]$ (mol L ⁻¹)	Δt (h)	Rate of Decomposition, (mol L ⁻¹ h ⁻¹)
0.00	1.000	-0.500	6.00	0.0833
6.00	0.500			
12.00	0.250	-0.250	6.00	0.0417
18.00	0.125			
24.00	0.0625	-0.062	6.00	0.010

The rate of decomposition of H_2O_2 in an aqueous solution decreases as the concentration of H_2O_2 decreases.

To obtain the tabulated results for this decomposition, the concentration of hydrogen peroxide was measured every 6 hours over the course of a day at a constant temperature of 40 °C. Reaction rates were computed for each time interval by dividing the change in concentration by the corresponding time increment, as shown here for the first 6-hour period:

Equation:

$$\frac{-\Delta[\text{H}_2\text{O}_2]}{\Delta t} = \frac{-(0.500 \text{ mol/L} - 1.000 \text{ mol/L})}{(6.00 \text{ h} - 0.00 \text{ h})} = 0.0833 \text{ mol L}^{-1} \text{ h}^{-1}$$

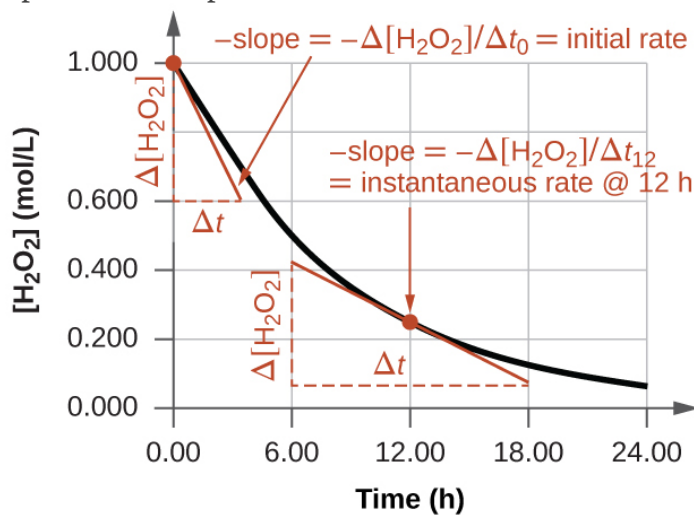
Notice that the reaction rates vary with time, decreasing as the reaction proceeds. Results for the last 6-hour period yield a reaction rate of:

Equation:

$$\frac{-\Delta[\text{H}_2\text{O}_2]}{\Delta t} = \frac{-(0.0625 \text{ mol/L} - 0.125 \text{ mol/L})}{(24.00 \text{ h} - 18.00 \text{ h})} = 0.010 \text{ mol L}^{-1} \text{ h}^{-1}$$

This behavior indicates the reaction continually slows with time. Using the concentrations at the beginning and end of a time period over which the reaction rate is changing results in the calculation of an **average rate** for the reaction over this time interval. At any specific time, the rate at which a reaction is proceeding is known as its **instantaneous rate**. The instantaneous rate of a reaction at “time zero,” when the reaction commences, is its **initial rate**. Consider the analogy of a car slowing down as it approaches a stop sign. The vehicle’s initial rate—analogue to the beginning of a chemical reaction—would be the speedometer reading at the moment the driver begins pressing the brakes (t_0). A few moments later, the instantaneous rate at a specific moment—call it t_1 —would be somewhat slower, as indicated by the speedometer reading at that point in time. As time passes, the instantaneous rate will continue to fall until it reaches zero, when the car (or reaction) stops. Unlike instantaneous speed, the car’s average speed is not indicated by the speedometer; but it can be calculated as the ratio of the distance traveled to the time required to bring the vehicle to a complete stop (Δt). Like the decelerating car, the average rate of a chemical reaction will fall somewhere between its initial and final rates.

The instantaneous rate of a reaction may be determined one of two ways. If experimental conditions permit the measurement of concentration changes over very short time intervals, then average rates computed as described earlier provide reasonably good approximations of instantaneous rates. Alternatively, a graphical procedure may be used that, in effect, yields the results that would be obtained if short time interval measurements were possible. In a plot of the concentration of hydrogen peroxide against time, the instantaneous rate of decomposition of H_2O_2 at any time t is given by the slope of a straight line that is tangent to the curve at that time ([link](#)). These tangent line slopes may be evaluated using calculus, but the procedure for doing so is beyond the scope of this chapter.



This graph shows a plot of concentration versus time for a 1.000 M solution of H_2O_2 . The rate

at any time is equal to the negative of the slope of a line tangent to the curve at that time. Tangents are shown at $t = 0$ h (“initial rate”) and at $t = 12$ h (“instantaneous rate” at 12 h).

Note:

Reaction Rates in Analysis: Test Strips for Urinalysis

Physicians often use disposable test strips to measure the amounts of various substances in a patient’s urine ([link](#)). These test strips contain various chemical reagents, embedded in small pads at various locations along the strip, which undergo changes in color upon exposure to sufficient concentrations of specific substances. The usage instructions for test strips often stress that proper read time is critical for optimal results. This emphasis on read time suggests that kinetic aspects of the chemical reactions occurring on the test strip are important considerations.

The test for urinary glucose relies on a two-step process represented by the chemical equations shown here:

Equation:

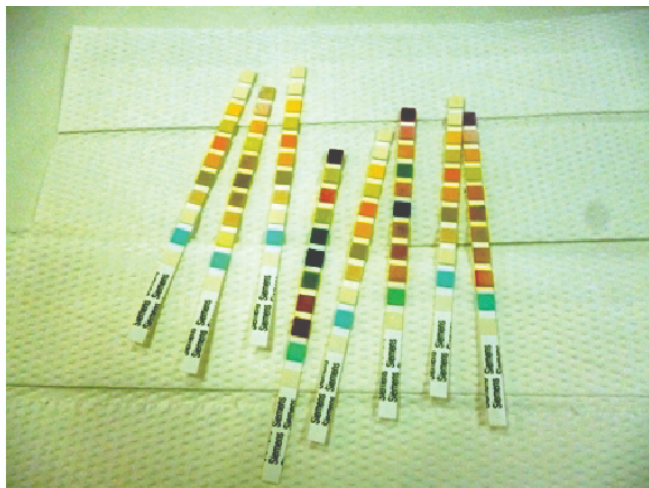


Equation:



The first equation depicts the oxidation of glucose in the urine to yield glucolactone and hydrogen peroxide. The hydrogen peroxide produced subsequently oxidizes colorless iodide ion to yield brown iodine, which may be visually detected. Some strips include an additional substance that reacts with iodine to produce a more distinct color change.

The two test reactions shown above are inherently very slow, but their rates are increased by special enzymes embedded in the test strip pad. This is an example of *catalysis*, a topic discussed later in this chapter. A typical glucose test strip for use with urine requires approximately 30 seconds for completion of the color-forming reactions. Reading the result too soon might lead one to conclude that the glucose concentration of the urine sample is lower than it actually is (a *false-negative* result). Waiting too long to assess the color change can lead to a *false positive* due to the slower (not catalyzed) oxidation of iodide ion by other substances found in urine.



Test strips are commonly used to detect the presence of specific substances in a person's urine. Many test strips have several pads containing various reagents to permit the detection of multiple substances on a single strip. (credit: Iqbal Osman)

Relative Rates of Reaction

The rate of a reaction may be expressed as the change in concentration of any reactant or product. For any given reaction, these rate expressions are all related simply to one another according to the reaction stoichiometry. The rate of the general reaction

Equation:



can be expressed in terms of the decrease in the concentration of A or the increase in the concentration of B. These two rate expressions are related by the stoichiometry of the reaction:

Equation:

$$\text{rate} = -\left(\frac{1}{a}\right)\left(\frac{\Delta A}{\Delta t}\right) = \left(\frac{1}{b}\right)\left(\frac{\Delta B}{\Delta t}\right)$$

the reaction represented by the following equation:

Equation:



The relation between the reaction rates expressed in terms of nitrogen production and ammonia consumption, for example, is:

Equation:

$$-\frac{\Delta \text{mol NH}_3}{\Delta t} \times \frac{1 \text{ mol N}_2}{2 \text{ mol NH}_3} = \frac{\Delta \text{mol N}_2}{\Delta t}$$

This may be represented in an abbreviated format by omitting the units of the stoichiometric factor:

Equation:

$$-\frac{1}{2} \frac{\Delta \text{mol NH}_3}{\Delta t} = \frac{\Delta \text{mol N}_2}{\Delta t}$$

Note that a negative sign has been included as a factor to account for the opposite signs of the two amount changes (the reactant amount is decreasing while the product amount is increasing). For homogeneous reactions, both the reactants and products are present in the same solution and thus occupy the same volume, so the molar amounts may be replaced with molar concentrations:

Equation:

$$-\frac{1}{2} \frac{\Delta [\text{NH}_3]}{\Delta t} = \frac{\Delta [\text{N}_2]}{\Delta t}$$

Similarly, the rate of formation of H_2 is three times the rate of formation of N_2 because three moles of H_2 are produced for each mole of N_2 produced.

Equation:

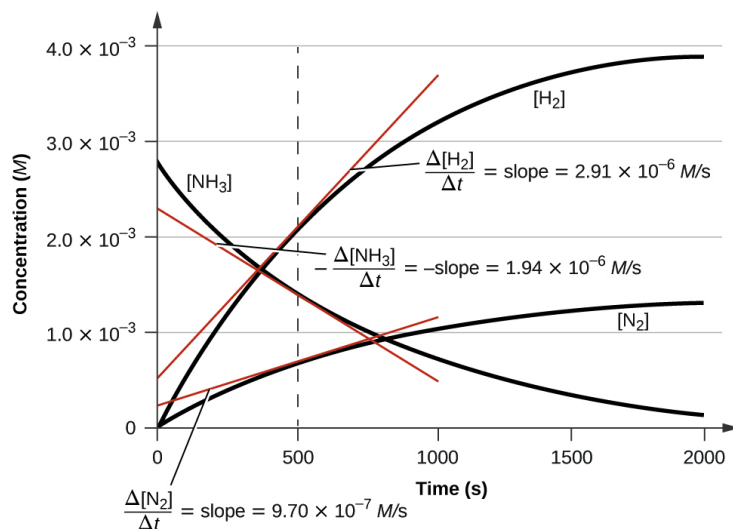
$$\frac{1}{3} \frac{\Delta [\text{H}_2]}{\Delta t} = \frac{\Delta [\text{N}_2]}{\Delta t}$$

[\[link\]](#) illustrates the change in concentrations over time for the decomposition of ammonia into nitrogen and hydrogen at 1100 °C. Slopes of the tangent lines at $t = 500 \text{ s}$ show that the instantaneous rates derived from all three species involved in the reaction

are related by their stoichiometric factors. The rate of hydrogen production, for example, is observed to be three times greater than that for nitrogen production:

Equation:

$$\frac{2.91 \times 10^{-6} \text{ M/s}}{9.70 \times 10^{-7} \text{ M/s}} \approx 3$$



Changes in concentrations of the reactant and products for the reaction

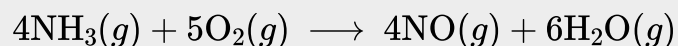
$2\text{NH}_3 \longrightarrow \text{N}_2 + 3\text{H}_2$. The rates of change of the three concentrations are related by the reaction stoichiometry, as shown by the different slopes of the tangents at $t = 500$ s.

Example:

Expressions for Relative Reaction Rates

The first step in the production of nitric acid is the combustion of ammonia:

Equation:



Write the equations that relate the rates of consumption of the reactants and the rates of formation of the products.

Solution

Considering the stoichiometry of this homogeneous reaction, the rates for the consumption of reactants and formation of products are:

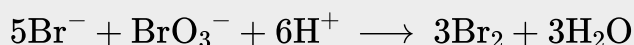
Equation:

$$-\frac{1}{4} \frac{\Delta[\text{NH}_3]}{\Delta t} = -\frac{1}{5} \frac{\Delta[\text{O}_2]}{\Delta t} = \frac{1}{4} \frac{\Delta[\text{NO}]}{\Delta t} = \frac{1}{6} \frac{\Delta[\text{H}_2\text{O}]}{\Delta t}$$

Check Your Learning

The rate of formation of Br_2 is $6.0 \times 10^{-6} \text{ mol/L/s}$ in a reaction described by the following net ionic equation:

Equation:



Write the equations that relate the rates of consumption of the reactants and the rates of formation of the products.

Note:

Answer:

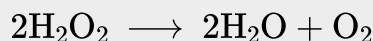
$$-\frac{1}{5} \frac{\Delta[\text{Br}^-]}{\Delta t} = -\frac{\Delta[\text{BrO}_3^-]}{\Delta t} = -\frac{1}{6} \frac{\Delta[\text{H}^+]}{\Delta t} = \frac{1}{3} \frac{\Delta[\text{Br}_2]}{\Delta t} = \frac{1}{3} \frac{\Delta[\text{H}_2\text{O}]}{\Delta t}$$

Example:

Reaction Rate Expressions for Decomposition of H_2O_2

The graph in [\[link\]](#) shows the rate of the decomposition of H_2O_2 over time:

Equation:



Based on these data, the instantaneous rate of decomposition of H_2O_2 at $t = 11.1 \text{ h}$ is determined to be

$3.20 \times 10^{-2} \text{ mol/L/h}$, that is:

Equation:

$$-\frac{\Delta[\text{H}_2\text{O}_2]}{\Delta t} = 3.20 \times 10^{-2} \text{ mol L}^{-1} \text{ h}^{-1}$$

What is the instantaneous rate of production of H_2O and O_2 ?

Solution

The reaction stoichiometry shows that

Equation:

$$-\frac{1}{2} \frac{\Delta[\text{H}_2\text{O}_2]}{\Delta t} = \frac{1}{2} \frac{\Delta[\text{H}_2\text{O}]}{\Delta t} = \frac{\Delta[\text{O}_2]}{\Delta t}$$

Therefore:

Equation:

$$\frac{1}{2} \times 3.20 \times 10^{-2} \text{ mol L}^{-1} \text{ h}^{-1} = \frac{\Delta[\text{O}_2]}{\Delta t}$$

and

Equation:

$$\frac{\Delta[\text{O}_2]}{\Delta t} = 1.60 \times 10^{-2} \text{ mol L}^{-1} \text{ h}^{-1}$$

Check Your Learning

If the rate of decomposition of ammonia, NH_3 , at 1150 K is $2.10 \times 10^{-6} \text{ mol/L/s}$, what is the rate of production of nitrogen and hydrogen?

Note:

Answer:

$1.05 \times 10^{-6} \text{ mol/L/s}$, N_2 and $3.15 \times 10^{-6} \text{ mol/L/s}$, H_2 .

Key Concepts and Summary

The rate of a reaction can be expressed either in terms of the decrease in the amount of a reactant or the increase in the amount of a product per unit time. Relations between different rate expressions for a given reaction are derived directly from the stoichiometric coefficients of the equation representing the reaction.

Key Equations

- relative reaction rates for $a\text{A} \longrightarrow b\text{B} = -\frac{1}{a} \frac{\Delta[\text{A}]}{\Delta t} = \frac{1}{b} \frac{\Delta[\text{B}]}{\Delta t}$

Chemistry End of Chapter Exercises

Exercise:

Problem:

What is the difference between average rate, initial rate, and instantaneous rate?

Solution:

The instantaneous rate is the rate of a reaction at any particular point in time, a period of time that is so short that the concentrations of reactants and products change by a negligible amount. The initial rate is the instantaneous rate of reaction as it starts (as product just begins to form). Average rate is the average of the instantaneous rates over a time period.

Exercise:

Problem:

Ozone decomposes to oxygen according to the equation $2\text{O}_3(g) \longrightarrow 3\text{O}_2(g)$. Write the equation that relates the rate expressions for this reaction in terms of the disappearance of O_3 and the formation of oxygen.

Exercise:

Problem:

In the nuclear industry, chlorine trifluoride is used to prepare uranium hexafluoride, a volatile compound of uranium used in the separation of uranium isotopes. Chlorine trifluoride is prepared by the reaction $\text{Cl}_2(g) + 3\text{F}_2(g) \longrightarrow 2\text{ClF}_3(g)$. Write the equation that relates the rate expressions for this reaction in terms of the disappearance of Cl_2 and F_2 and the formation of ClF_3 .

Solution:

$$\text{rate} = +\frac{1}{2} \frac{\Delta[\text{ClF}_3]}{\Delta t} = -\frac{\Delta[\text{Cl}_2]}{\Delta t} = -\frac{1}{3} \frac{\Delta[\text{F}_2]}{\Delta t}$$

Exercise:

Problem:

A study of the rate of dimerization of C_4H_6 gave the data shown in the table:
 $2\text{C}_4\text{H}_6 \longrightarrow \text{C}_8\text{H}_{12}$

Time (s)	0	1600	3200	4800	6200
$[\text{C}_4\text{H}_6]$ (M)	1.00×10^{-2}	5.04×10^{-3}	3.37×10^{-3}	2.53×10^{-3}	2.08×10^{-3}

- (a) Determine the average rate of dimerization between 0 s and 1600 s, and between 1600 s and 3200 s.
- (b) Estimate the instantaneous rate of dimerization at 3200 s from a graph of time versus $[\text{C}_4\text{H}_6]$. What are the units of this rate?
- (c) Determine the average rate of formation of C_8H_{12} at 1600 s and the instantaneous rate of formation at 3200 s from the rates found in parts (a) and (b).

Exercise:

Problem:

A study of the rate of the reaction represented as $2A \longrightarrow B$ gave the following data:

Time (s)	0.0	5.0	10.0	15.0	20.0	25.0	35.0
$[A]$ (M)	1.00	0.775	0.625	0.465	0.360	0.285	0.230

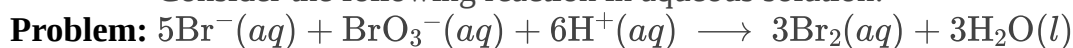
- (a) Determine the average rate of disappearance of A between 0.0 s and 10.0 s, and between 10.0 s and 20.0 s.
- (b) Estimate the instantaneous rate of disappearance of A at 15.0 s from a graph of time versus $[A]$. What are the units of this rate?
- (c) Use the rates found in parts (a) and (b) to determine the average rate of formation of B between 0.00 s and 10.0 s, and the instantaneous rate of formation of B at 15.0 s.

Solution:

(a) average rate, $0 - 10 \text{ s} = 0.0375 \text{ mol L}^{-1} \text{ s}^{-1}$; average rate, $10 - 20 \text{ s} = 0.0265 \text{ mol L}^{-1} \text{ s}^{-1}$; (b) instantaneous rate, $15 \text{ s} = 0.023 \text{ mol L}^{-1} \text{ s}^{-1}$; (c) average rate for B formation $= 0.0188 \text{ mol L}^{-1} \text{ s}^{-1}$; instantaneous rate for B formation $= 0.012 \text{ mol L}^{-1} \text{ s}^{-1}$

Exercise:

Consider the following reaction in aqueous solution:



If the rate of disappearance of $\text{Br}^-(aq)$ at a particular moment during the reaction is $3.5 \times 10^{-4} \text{ mol L}^{-1} \text{ s}^{-1}$, what is the rate of appearance of $\text{Br}_2(aq)$ at that moment?

Glossary**average rate**

rate of a chemical reaction computed as the ratio of a measured change in amount or concentration of substance to the time interval over which the change occurred

initial rate

instantaneous rate of a chemical reaction at $t = 0 \text{ s}$ (immediately after the reaction has begun)

instantaneous rate

rate of a chemical reaction at any instant in time, determined by the slope of the line tangential to a graph of concentration as a function of time

rate of reaction

measure of the speed at which a chemical reaction takes place

rate expression

mathematical representation defining reaction rate as change in amount, concentration, or pressure of reactant or product species per unit time

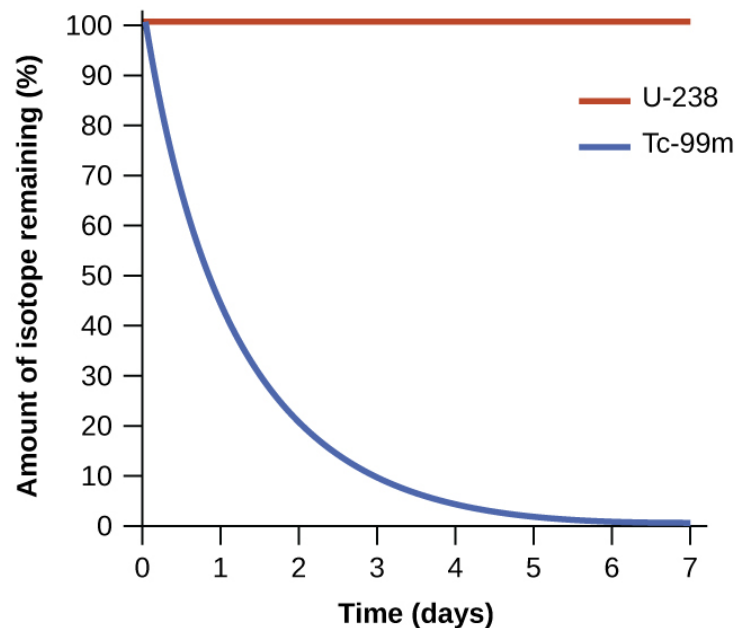
Spontaneity

By the end of this section, you will be able to:

- Distinguish between spontaneous and nonspontaneous processes
- Describe the dispersal of matter and energy that accompanies certain spontaneous processes

Processes have a natural tendency to occur in one direction under a given set of conditions. Water will naturally flow downhill, but uphill flow requires outside intervention such as the use of a pump. Iron exposed to the earth's atmosphere will corrode, but rust is not converted to iron without intentional chemical treatment. A **spontaneous process** is one that occurs naturally under certain conditions. A **nonspontaneous process**, on the other hand, will not take place unless it is “driven” by the continual input of energy from an external source. A process that is spontaneous in one direction under a particular set of conditions is nonspontaneous in the reverse direction. At room temperature and typical atmospheric pressure, for example, ice will spontaneously melt, but water will not spontaneously freeze.

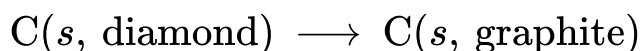
The spontaneity of a process is *not* correlated to the speed of the process. A spontaneous change may be so rapid that it is essentially instantaneous or so slow that it cannot be observed over any practical period of time. To illustrate this concept, consider the decay of radioactive isotopes, a topic more thoroughly treated in the chapter on nuclear chemistry. Radioactive decay is by definition a spontaneous process in which the nuclei of unstable isotopes emit radiation as they are converted to more stable nuclei. All the decay processes occur spontaneously, but the rates at which different isotopes decay vary widely. Technetium-99m is a popular radioisotope for medical imaging studies that undergoes relatively rapid decay and exhibits a half-life of about six hours. Uranium-238 is the most abundant isotope of uranium, and its decay occurs much more slowly, exhibiting a half-life of more than four billion years ([\[link\]](#)).



Both U-238 and Tc-99m undergo spontaneous radioactive decay, but at drastically different rates. Over the course of one week, essentially all of a Tc-99m sample and none of a U-238 sample will have decayed.

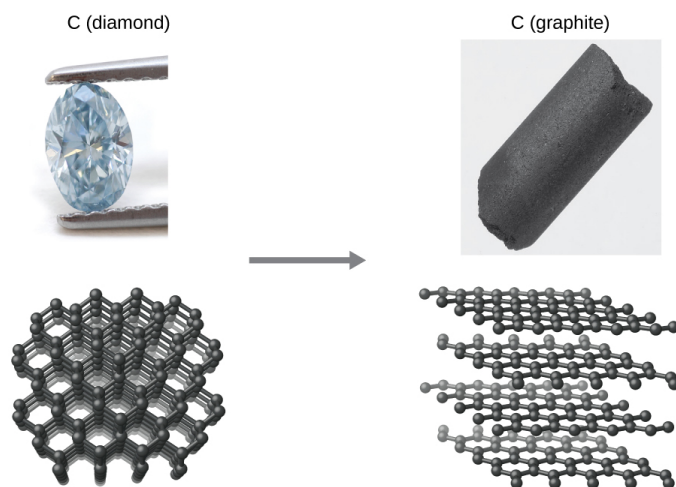
As another example, consider the conversion of diamond into graphite ([link](#)).

Equation:



The phase diagram for carbon indicates that graphite is the stable form of this element under ambient atmospheric pressure, while diamond is the stable allotrope at very high pressures, such as those present during its geologic formation. Thermodynamic calculations of the sort described in the last section of this chapter indicate that the conversion of diamond to graphite at ambient pressure occurs spontaneously, yet diamonds are

observed to exist, and persist, under these conditions. Though the process is spontaneous under typical ambient conditions, its rate is extremely slow; so, for all practical purposes diamonds are indeed “forever.” Situations such as these emphasize the important distinction between the thermodynamic and the kinetic aspects of a process. In this particular case, diamonds are said to be *thermodynamically unstable* but *kinetically stable* under ambient conditions.



The conversion of carbon from the diamond allotrope to the graphite allotrope is spontaneous at ambient pressure, but its rate is immeasurably slow at low to moderate temperatures. This process is known as *graphitization*, and its rate can be increased to easily measurable values at temperatures in the 1000–2000 K range. (credit "diamond" photo: modification of work by "Fancy Diamonds"/Flickr; credit "graphite" photo: modification of work by images-of-elements.com/carbon.php)

Dispersal of Matter and Energy

Extending the discussion of thermodynamic concepts toward the objective of predicting spontaneity, consider now an isolated system consisting of two flasks connected with a closed valve. Initially there is an ideal gas in one flask and the other flask is empty ($P = 0$). ([link](#)). When the valve is opened, the gas spontaneously expands to fill both flasks equally. Recalling the definition of pressure-volume work from the chapter on thermochemistry, note that no work has been done because the pressure in a vacuum is zero.

Equation:

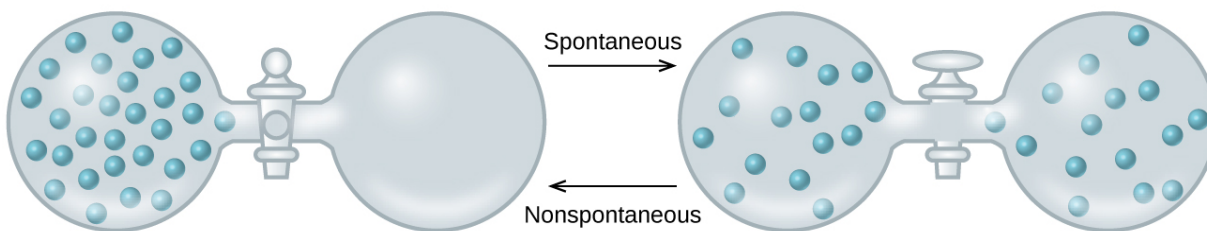
$$w = -P\Delta V = 0 \quad (P = 0 \text{ in a vacuum})$$

Note as well that since the system is isolated, no heat has been exchanged with the surroundings ($q = 0$). The *first law of thermodynamics* confirms that there has been no change in the system's internal energy as a result of this process.

Equation:

$$\Delta U = q + w = 0 + 0 = 0$$

The spontaneity of this process is therefore not a consequence of any change in energy that accompanies the process. Instead, the driving force appears to be related to the *greater, more uniform dispersal of matter* that results when the gas is allowed to expand. Initially, the system was comprised of one flask containing matter and another flask containing nothing. After the spontaneous expansion took place, the matter was distributed both more widely (occupying twice its original volume) and more uniformly (present in equal amounts in each flask).



An isolated system consists of an ideal gas in one flask that is connected by a closed valve to a second flask containing a vacuum. Once the valve is opened, the gas spontaneously becomes evenly distributed between the flasks.

Now consider two objects at different temperatures: object X at temperature T_X and object Y at temperature T_Y , with $T_X > T_Y$ ([link](#)). When these objects come into contact, heat spontaneously flows from the hotter object (X) to the colder one (Y). This corresponds to a loss of thermal energy by X and a gain of thermal energy by Y.

Equation:

$$q_X < 0 \quad \text{and} \quad q_Y = -q_X > 0$$

From the perspective of this two-object system, there was no net gain or loss of thermal energy, rather the available thermal energy was redistributed among the two objects. This spontaneous process resulted in a *more uniform dispersal of energy*.



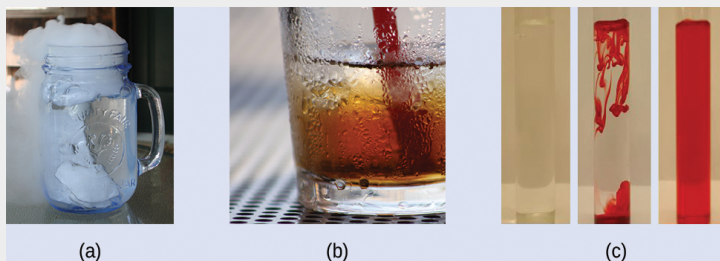
When two objects at different temperatures come in contact, heat spontaneously flows from the hotter to the colder object.

As illustrated by the two processes described, an important factor in determining the spontaneity of a process is the extent to which it changes the dispersal or distribution of matter and/or energy. In each case, a spontaneous process took place that resulted in a more uniform distribution of matter or energy.

Example:**Redistribution of Matter during a Spontaneous Process**

Describe how matter is redistributed when the following spontaneous processes take place:

- (a) A solid sublimates.
- (b) A gas condenses.
- (c) A drop of food coloring added to a glass of water forms a solution with uniform color.

Solution

(credit a: modification of work by Jenny Downing; credit b: modification of work by “Fuzzy Gerdes”/Flickr; credit c: modification of work by Paul A. Flowers)

(a) Sublimation is the conversion of a solid (relatively high density) to a gas (much lesser density). This process yields a much greater dispersal of matter, since the molecules will occupy a much greater volume after the solid-to-gas transition.

(b) Condensation is the conversion of a gas (relatively low density) to a liquid (much greater density). This process yields a much lesser dispersal of matter, since the molecules will occupy a much lesser volume after the gas-to-liquid transition.

(c) The process in question is diffusion. This process yields a more uniform dispersal of matter, since the initial state of the system involves two regions of different dye concentrations (high in the drop of dye, zero in the water), and the final state of the system contains a single dye concentration throughout.

Check Your Learning

Describe how energy is redistributed when a spoon at room temperature is placed in a cup of hot coffee.

Note:

Answer:

Heat will spontaneously flow from the hotter object (coffee) to the colder object (spoon), resulting in a more uniform distribution of thermal energy as the spoon warms and the coffee cools.

Key Concepts and Summary

Chemical and physical processes have a natural tendency to occur in one direction under certain conditions. A spontaneous process occurs without the need for a continual input of energy from some external source, while a nonspontaneous process requires such. Systems undergoing a spontaneous process may or may not experience a gain or loss of energy, but they will experience a change in the way matter and/or energy is distributed within the system.

Chemistry End of Chapter Exercises

Exercise:

Problem: What is a spontaneous reaction?

Solution:

A reaction has a natural tendency to occur and takes place without the continual input of energy from an external source.

Exercise:

Problem: What is a nonspontaneous reaction?

Exercise:

Problem:

Indicate whether the following processes are spontaneous or nonspontaneous.

- (a) Liquid water freezing at a temperature below its freezing point
 - (b) Liquid water freezing at a temperature above its freezing point
 - (c) The combustion of gasoline
 - (d) A ball thrown into the air
 - (e) A raindrop falling to the ground
 - (f) Iron rusting in a moist atmosphere
-

Solution:

(a) spontaneous; (b) nonspontaneous; (c) spontaneous; (d) nonspontaneous; (e) spontaneous; (f) spontaneous

Exercise:

Problem:

A helium-filled balloon spontaneously deflates overnight as He atoms diffuse through the wall of the balloon. Describe the redistribution of matter and/or energy that accompanies this process.

Exercise:

Problem:

Many plastic materials are organic polymers that contain carbon and hydrogen. The oxidation of these plastics in air to form carbon dioxide and water is a spontaneous process; however, plastic materials tend to persist in the environment. Explain.

Solution:

Although the oxidation of plastics is spontaneous, the rate of oxidation is very slow. Plastics are therefore kinetically stable and do not decompose appreciably even over relatively long periods of time.

Glossary

nonspontaneous process

process that requires continual input of energy from an external source

spontaneous change

process that takes place without a continuous input of energy from an external source

Entropy

By the end of this section, you will be able to:

- Define entropy
- Explain the relationship between entropy and the number of microstates
- Predict the sign of the entropy change for chemical and physical processes

In 1824, at the age of 28, Nicolas Léonard Sadi Carnot ([\[link\]](#)) published the results of an extensive study regarding the efficiency of steam heat engines. A later review of Carnot's findings by Rudolf Clausius introduced a new thermodynamic property that relates the spontaneous heat flow accompanying a process to the temperature at which the process takes place. This new property was expressed as the ratio of the *reversible* heat (q_{rev}) and the kelvin temperature (T). In thermodynamics, a **reversible process** is one that takes place at such a slow rate that it is always at equilibrium and its direction can be changed (it can be “reversed”) by an infinitesimally small change in some condition. Note that the idea of a reversible process is a formalism required to support the development of various thermodynamic concepts; no real processes are truly reversible, rather they are classified as *irreversible*.



(a)



(b)

(a) Nicholas Léonard Sadi Carnot's research into steam-powered machinery and (b) Rudolf Clausius's later study of those findings led to groundbreaking discoveries about spontaneous heat flow processes.

Similar to other thermodynamic properties, this new quantity is a state function, so its change depends only upon the initial and final states of a system. In 1865, Clausius named this property **entropy (S)** and defined its change for any process as the following:

Equation:

$$\Delta S = \frac{q_{\text{rev}}}{T}$$

The entropy change for a real, irreversible process is then equal to that for the theoretical reversible process that involves the same initial and final states.

Entropy and Microstates

Following the work of Carnot and Clausius, Ludwig Boltzmann developed a molecular-scale statistical model that related the entropy of a system to the *number of microstates* (W) possible for the system. A **microstate** is a specific configuration of all the locations and energies of the atoms or molecules that make up a system. The relation between a system's entropy and the number of possible microstates is

Equation:

$$S = k \ln W$$

where k is the Boltzmann constant, 1.38×10^{-23} J/K.

As for other state functions, the change in entropy for a process is the difference between its final (S_f) and initial (S_i) values:

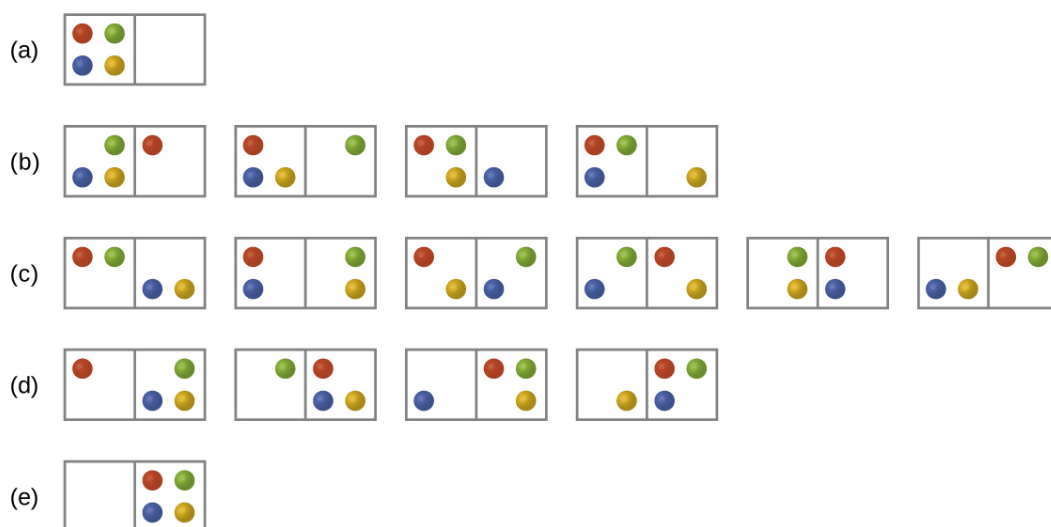
Equation:

$$\Delta S = S_f - S_i = k \ln W_f - k \ln W_i = k \ln \frac{W_f}{W_i}$$

For processes involving an increase in the number of microstates, $W_f > W_i$, the entropy of the system increases and $\Delta S > 0$. Conversely, processes that reduce

the number of microstates, $W_f < W_i$, yield a decrease in system entropy, $\Delta S < 0$. This molecular-scale interpretation of entropy provides a link to the probability that a process will occur as illustrated in the next paragraphs.

Consider the general case of a system comprised of N particles distributed among n boxes. The number of microstates possible for such a system is n^N . For example, distributing four particles among two boxes will result in $2^4 = 16$ different microstates as illustrated in [\[link\]](#). Microstates with equivalent particle arrangements (not considering individual particle identities) are grouped together and are called *distributions*. The probability that a system will exist with its components in a given distribution is proportional to the number of microstates within the distribution. Since entropy increases logarithmically with the number of microstates, *the most probable distribution is therefore the one of greatest entropy*.



The sixteen microstates associated with placing four particles in two boxes are shown. The microstates are collected into five distributions—(a), (b), (c), (d), and (e)—based on the numbers of particles in each box.

For this system, the most probable configuration is one of the six microstates associated with distribution (c) where the particles are evenly distributed between the boxes, that is, a configuration of two particles in each box. The

probability of finding the system in this configuration is $\frac{6}{16}$ or $\frac{3}{8}$. The least probable configuration of the system is one in which all four particles are in one box, corresponding to distributions (a) and (e), each with a probability of $\frac{1}{16}$. The probability of finding all particles in only one box (either the left box or right box) is then $\left(\frac{1}{16} + \frac{1}{16}\right) = \frac{2}{16}$ or $\frac{1}{8}$.

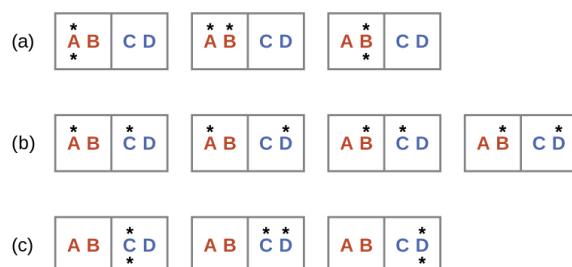
As you add more particles to the system, the number of possible microstates increases exponentially (2^N). A macroscopic (laboratory-sized) system would typically consist of moles of particles ($N \sim 10^{23}$), and the corresponding number of microstates would be staggeringly huge. Regardless of the number of particles in the system, however, the distributions in which roughly equal numbers of particles are found in each box are always the most probable configurations.

This matter dispersal model of entropy is often described qualitatively in terms of the *disorder* of the system. By this description, microstates in which all the particles are in a single box are the most ordered, thus possessing the least entropy. Microstates in which the particles are more evenly distributed among the boxes are more disordered, possessing greater entropy.

The previous description of an ideal gas expanding into a vacuum ([\[link\]](#)) is a macroscopic example of this particle-in-a-box model. For this system, the most probable distribution is confirmed to be the one in which the matter is most uniformly dispersed or distributed between the two flasks. Initially, the gas molecules are confined to just one of the two flasks. Opening the valve between the flasks increases the volume available to the gas molecules and, correspondingly, the number of microstates possible for the system. Since $W_f > W_i$, the expansion process involves an increase in entropy ($\Delta S > 0$) and is spontaneous.

A similar approach may be used to describe the spontaneous flow of heat. Consider a system consisting of two objects, each containing two particles, and two units of thermal energy (represented as “*”) in [\[link\]](#). The hot object is comprised of particles **A** and **B** and initially contains both energy units. The cold object is comprised of particles **C** and **D**, which initially has no energy units. Distribution (a) shows the three microstates possible for the initial state of the system, with both units of energy contained within the hot object. If one of the two energy units is transferred, the result is distribution (b) consisting of

four microstates. If both energy units are transferred, the result is distribution (c) consisting of three microstates. Thus, we may describe this system by a total of ten microstates. The probability that the heat does not flow when the two objects are brought into contact, that is, that the system remains in distribution (a), is $\frac{3}{10}$. More likely is the flow of heat to yield one of the other two distributions, the combined probability being $\frac{7}{10}$. The most likely result is the flow of heat to yield the uniform dispersal of energy represented by distribution (b), the probability of this configuration being $\frac{4}{10}$. This supports the common observation that placing hot and cold objects in contact results in spontaneous heat flow that ultimately equalizes the objects' temperatures. And, again, this spontaneous process is also characterized by an increase in system entropy.

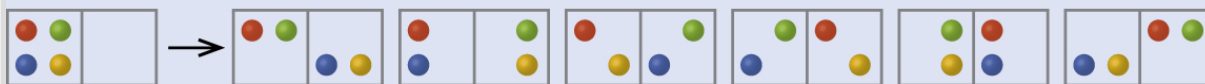


This shows a microstate model describing the flow of heat from a hot object to a cold object. (a) Before the heat flow occurs, the object comprised of particles **A** and **B** contains both units of energy and as represented by a distribution of three microstates. (b) If the heat flow results in an even dispersal of energy (one energy unit transferred), a distribution of four microstates results. (c) If both energy units are transferred, the resulting distribution has three microstates.

Example:

Determination of ΔS

Calculate the change in entropy for the process depicted below.



Solution

The initial number of microstates is one, the final six:

Equation:

$$\Delta S = k \ln \frac{W_c}{W_a} = 1.38 \times 10^{-23} \text{ J/K} \times \ln \frac{6}{1} = 2.47 \times 10^{-23} \text{ J/K}$$

The sign of this result is consistent with expectation; since there are more microstates possible for the final state than for the initial state, the change in entropy should be positive.

Check Your Learning

Consider the system shown in [\[link\]](#). What is the change in entropy for the process where *all* the energy is transferred from the hot object (**AB**) to the cold object (**CD**)?

Note:

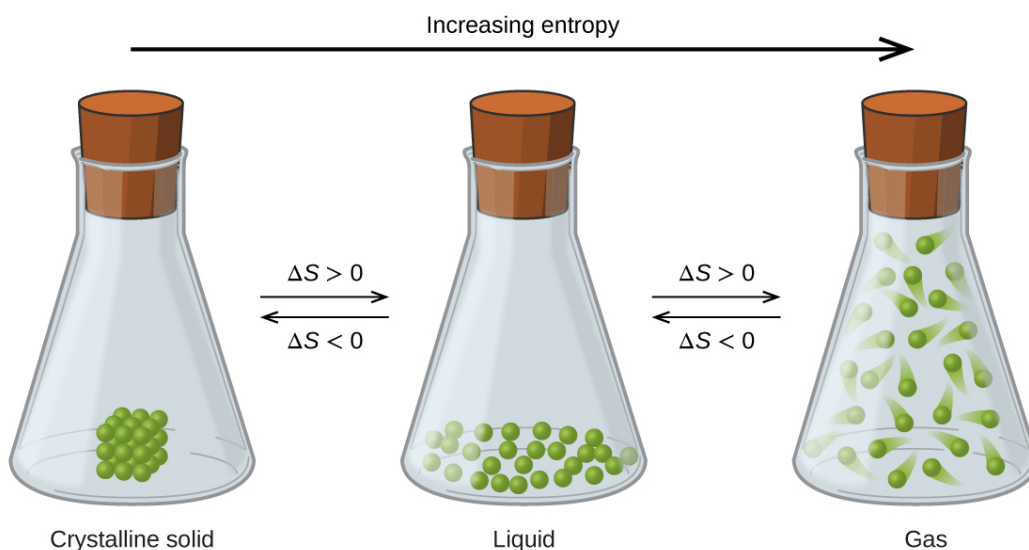
Answer:

0 J/K

Predicting the Sign of ΔS

The relationships between entropy, microstates, and matter/energy dispersal described previously allow us to make generalizations regarding the relative entropies of substances and to predict the sign of entropy changes for chemical and physical processes. Consider the phase changes illustrated in [\[link\]](#). In the solid phase, the atoms or molecules are restricted to nearly fixed positions with

respect to each other and are capable of only modest oscillations about these positions. With essentially fixed locations for the system's component particles, the number of microstates is relatively small. In the liquid phase, the atoms or molecules are free to move over and around each other, though they remain in relatively close proximity to one another. This increased freedom of motion results in a greater variation in possible particle locations, so the number of microstates is correspondingly greater than for the solid. As a result, $S_{\text{liquid}} > S_{\text{solid}}$ and the process of converting a substance from solid to liquid (melting) is characterized by an increase in entropy, $\Delta S > 0$. By the same logic, the reciprocal process (freezing) exhibits a decrease in entropy, $\Delta S < 0$.

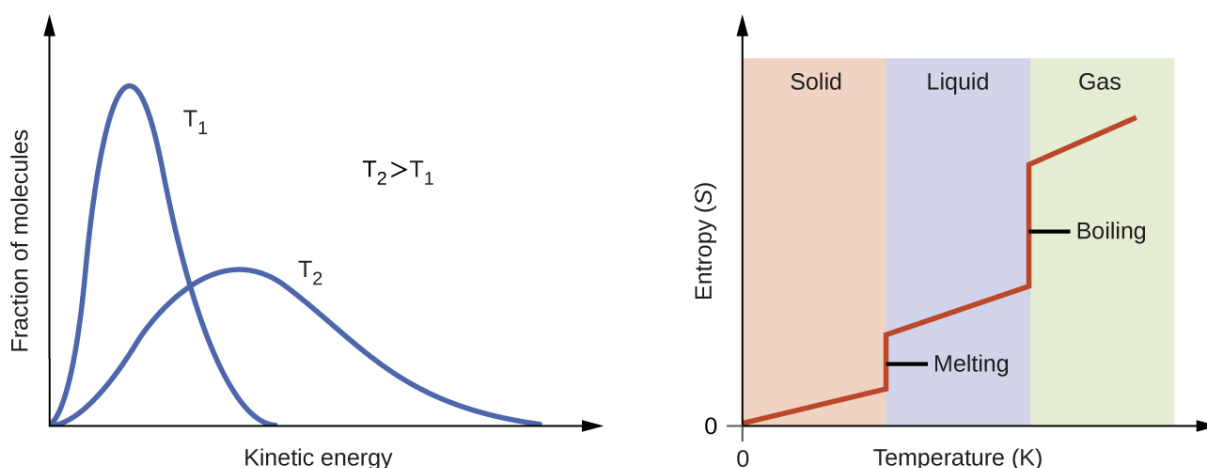


The entropy of a substance increases ($\Delta S > 0$) as it transforms from a relatively ordered solid, to a less-ordered liquid, and then to a still less-ordered gas. The entropy decreases ($\Delta S < 0$) as the substance transforms from a gas to a liquid and then to a solid.

Now consider the gaseous phase, in which a given number of atoms or molecules occupy a *much* greater volume than in the liquid phase. Each atom or molecule can be found in many more locations, corresponding to a much greater number of microstates. Consequently, for any substance, $S_{\text{gas}} > S_{\text{liquid}} >$

S_{solid} , and the processes of vaporization and sublimation likewise involve increases in entropy, $\Delta S > 0$. Likewise, the reciprocal phase transitions, condensation and deposition, involve decreases in entropy, $\Delta S < 0$.

According to kinetic-molecular theory, the temperature of a substance is proportional to the average kinetic energy of its particles. Raising the temperature of a substance will result in more extensive vibrations of the particles in solids and more rapid translations of the particles in liquids and gases. At higher temperatures, the distribution of kinetic energies among the atoms or molecules of the substance is also broader (more dispersed) than at lower temperatures. Thus, the entropy for any substance increases with temperature ([link](#)).



Entropy increases as the temperature of a substance is raised, which corresponds to the greater spread of kinetic energies. When a substance undergoes a phase transition, its entropy changes significantly.

Note:

Try this [simulator](#) with interactive visualization of the dependence of particle location and freedom of motion on physical state and temperature.

The entropy of a substance is influenced by the structure of the particles (atoms or molecules) that comprise the substance. With regard to atomic substances, heavier atoms possess greater entropy at a given temperature than lighter atoms, which is a consequence of the relation between a particle's mass and the spacing of quantized translational energy levels (a topic beyond the scope of this text). For molecules, greater numbers of atoms increase the number of ways in which the molecules can vibrate and thus the number of possible microstates and the entropy of the system.

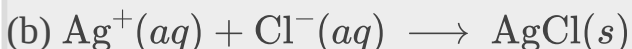
Finally, variations in the types of particles affects the entropy of a system. Compared to a pure substance, in which all particles are identical, the entropy of a mixture of two or more different particle types is greater. This is because of the additional orientations and interactions that are possible in a system comprised of nonidentical components. For example, when a solid dissolves in a liquid, the particles of the solid experience both a greater freedom of motion and additional interactions with the solvent particles. This corresponds to a more uniform dispersal of matter and energy and a greater number of microstates. The process of dissolution therefore involves an increase in entropy, $\Delta S > 0$.

Considering the various factors that affect entropy allows us to make informed predictions of the sign of ΔS for various chemical and physical processes as illustrated in [\[link\]](#).

Example:**Predicting the Sign of ΔS**

Predict the sign of the entropy change for the following processes. Indicate the reason for each of your predictions.

(a) One mole liquid water at room temperature $\longrightarrow$ one mole liquid water at 50 °C

**Solution**

(a) positive, temperature increases

(b) negative, reduction in the number of ions (particles) in solution, decreased dispersal of matter

- (c) negative, net decrease in the amount of gaseous species
- (d) positive, phase transition from solid to liquid, net increase in dispersal of matter

Check Your Learning

Predict the sign of the entropy change for the following processes. Give a reason for your prediction.

- (a) $\text{NaNO}_3(s) \longrightarrow \text{Na}^+(aq) + \text{NO}_3^-(aq)$
- (b) the freezing of liquid water
- (c) $\text{CO}_2(s) \longrightarrow \text{CO}_2(g)$
- (d) $\text{CaCO}(s) \longrightarrow \text{CaO}(s) + \text{CO}_2(g)$

Note:

Answer:

(a) Positive; The solid dissolves to give an increase of mobile ions in solution. (b) Negative; The liquid becomes a more ordered solid. (c) Positive; The relatively ordered solid becomes a gas. (d) Positive; There is a net increase in the amount of gaseous species.

Key Concepts and Summary

Entropy (S) is a state function that can be related to the number of microstates for a system (the number of ways the system can be arranged) and to the ratio of reversible heat to kelvin temperature. It may be interpreted as a measure of the dispersal or distribution of matter and/or energy in a system, and it is often described as representing the “disorder” of the system.

For a given substance, entropy depends on phase with $S_{\text{solid}} < S_{\text{liquid}} < S_{\text{gas}}$. For different substances in the same physical state at a given temperature, entropy is typically greater for heavier atoms or more complex molecules. Entropy increases when a system is heated and when solutions form. Using these guidelines, the sign of entropy changes for some chemical reactions and physical changes may be reliably predicted.

Key Equations

- $\Delta S = \frac{q_{\text{rev}}}{T}$
- $S = k \ln W$
- $\Delta S = k \ln \frac{W_f}{W_i}$

Chemistry End of Chapter Exercises

Exercise:

Problem:

In [\[link\]](#) all possible distributions and microstates are shown for four different particles shared between two boxes. Determine the entropy change, ΔS , if the particles are initially evenly distributed between the two boxes, but upon redistribution all end up in Box (b).

Exercise:

Problem:

In [\[link\]](#) all of the possible distributions and microstates are shown for four different particles shared between two boxes. Determine the entropy change, ΔS , for the system when it is converted from distribution (b) to distribution (d).

Solution:

There are four initial microstates and four final microstates.

$$\Delta S = k \ln \frac{W_f}{W_i} = 1.38 \times 10^{-23} \text{ J/K} \times \ln \frac{4}{4} = 0$$

Exercise:

Problem:

How does the process described in the previous item relate to the system shown in [\[link\]](#)?

Exercise:

Problem:

Consider a system similar to the one in [\[link\]](#), except that it contains six particles instead of four. What is the probability of having all the particles in only one of the two boxes in the case? Compare this with the similar probability for the system of four particles that we have derived to be equal to $\frac{1}{8}$. What does this comparison tell us about even larger systems?

Solution:

The probability for all the particles to be on one side is $\frac{1}{32}$. This probability is noticeably lower than the $\frac{1}{8}$ result for the four-particle system. The conclusion we can make is that the probability for all the particles to stay in only one part of the system will decrease rapidly as the number of particles increases, and, for instance, the probability for all molecules of gas to gather in only one side of a room at room temperature and pressure is negligible since the number of gas molecules in the room is very large.

Exercise:**Problem:**

Consider the system shown in [\[link\]](#). What is the change in entropy for the process where the energy is initially associated only with particle A, but in the final state the energy is distributed between two different particles?

Exercise:**Problem:**

Consider the system shown in [\[link\]](#). What is the change in entropy for the process where the energy is initially associated with particles A and B, and the energy is distributed between two particles in different boxes (one in A-B, the other in C-D)?

Solution:

There is only one initial state. For the final state, the energy can be contained in pairs A-C, A-D, B-C, or B-D. Thus, there are four final possible states.

$$\Delta S = k \ln \left(\frac{W_f}{W_i} \right) = 1.38 \times 10^{-23} \text{ J/K} \times \ln \left(\frac{4}{1} \right) = 1.91 \times 10^{-23} \text{ J/K}$$

Exercise:

Problem:

Arrange the following sets of systems in order of increasing entropy. Assume one mole of each substance and the same temperature for each member of a set.

(a) $\text{H}_2(g)$, $\text{HBrO}_4(g)$, $\text{HBr}(g)$

(b) $\text{H}_2\text{O}(l)$, $\text{H}_2\text{O}(g)$, $\text{H}_2\text{O}(s)$

(c) $\text{He}(g)$, $\text{Cl}_2(g)$, $\text{P}_4(g)$

Exercise:

Problem:

At room temperature, the entropy of the halogens increases from I_2 to Br_2 to Cl_2 . Explain.

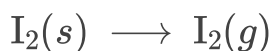
Solution:

The masses of these molecules would suggest the opposite trend in their entropies. The observed trend is a result of the more significant variation of entropy with a physical state. At room temperature, I_2 is a solid, Br_2 is a liquid, and Cl_2 is a gas.

Exercise:

Problem:

Consider two processes: sublimation of $\text{I}_2(s)$ and melting of $\text{I}_2(s)$ (Note: the latter process can occur at the same temperature but somewhat higher pressure).





Is ΔS positive or negative in these processes? In which of the processes will the magnitude of the entropy change be greater?

Exercise:

Problem:

Indicate which substance in the given pairs has the higher entropy value. Explain your choices.

(a) $\text{C}_2\text{H}_5\text{OH}(l)$ or $\text{C}_3\text{H}_7\text{OH}(l)$

(b) $\text{C}_2\text{H}_5\text{OH}(l)$ or $\text{C}_2\text{H}_5\text{OH}(g)$

(c) $2\text{H}(g)$ or $\text{H}(g)$

Solution:

(a) $\text{C}_3\text{H}_7\text{OH}(l)$ as it is a larger molecule (more complex and more massive), and so more microstates describing its motions are available at any given temperature. (b) $\text{C}_2\text{H}_5\text{OH}(g)$ as it is in the gaseous state. (c) $2\text{H}(g)$, since entropy is an extensive property, and so two H atoms (or two moles of H atoms) possess twice as much entropy as one atom (or one mole of atoms).

Exercise:

Problem:

Predict the sign of the entropy change for the following processes.

(a) An ice cube is warmed to near its melting point.

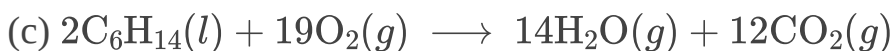
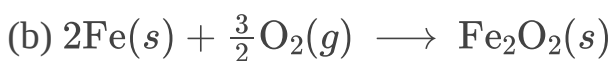
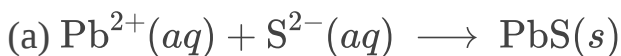
(b) Exhaled breath forms fog on a cold morning.

(c) Snow melts.

Exercise:

Problem:

Predict the sign of the entropy change for the following processes. Give a reason for your prediction.

**Solution:**

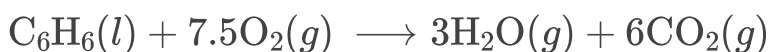
(a) Negative. The relatively ordered solid precipitating decreases the number of mobile ions in solution. (b) Negative. There is a net loss of three moles of gas from reactants to products. (c) Positive. There is a net increase of seven moles of gas from reactants to products.

Exercise:**Problem:**

Write the balanced chemical equation for the combustion of methane, $\text{CH}_4(\text{g})$, to give carbon dioxide and water vapor. Explain why it is difficult to predict whether ΔS is positive or negative for this chemical reaction.

Exercise:**Problem:**

Write the balanced chemical equation for the combustion of benzene, $\text{C}_6\text{H}_6(\text{l})$, to give carbon dioxide and water vapor. Would you expect ΔS to be positive or negative in this process?

Solution:

There are 7.5 moles of gas initially, and $3 + 6 = 9$ moles of gas in the end. Therefore, it is likely that the entropy increases as a result of this reaction, and ΔS is positive.

Glossary

entropy (S)

state function that is a measure of the matter and/or energy dispersal within a system, determined by the number of system microstates; often described as a measure of the disorder of the system

microstate

possible configuration or arrangement of matter and energy within a system

reversible process

process that takes place so slowly as to be capable of reversing direction in response to an infinitesimally small change in conditions; hypothetical construct that can only be approximated by real processes

The Second and Third Laws of Thermodynamics

By the end of this section, you will be able to:

- State and explain the second and third laws of thermodynamics
- Calculate entropy changes for phase transitions and chemical reactions under standard conditions

The Second Law of Thermodynamics

In the quest to identify a property that may reliably predict the spontaneity of a process, a promising candidate has been identified: entropy. Processes that involve an increase in entropy *of the system* ($\Delta S > 0$) are very often spontaneous; however, examples to the contrary are plentiful. By expanding consideration of entropy changes to include *the surroundings*, we may reach a significant conclusion regarding the relation between this property and spontaneity. In thermodynamic models, the system and surroundings comprise everything, that is, the universe, and so the following is true:

Equation:

$$\Delta S_{\text{univ}} = \Delta S_{\text{sys}} + \Delta S_{\text{surr}}$$

To illustrate this relation, consider again the process of heat flow between two objects, one identified as the system and the other as the surroundings. There are three possibilities for such a process:

1. The objects are at different temperatures, and heat flows from the hotter to the cooler object. *This is always observed to occur spontaneously.* Designating the hotter object as the system and invoking the definition of entropy yields the following:

Equation:

$$\Delta S_{\text{sys}} = \frac{-q_{\text{rev}}}{T_{\text{sys}}} \quad \text{and} \quad \Delta S_{\text{surr}} = \frac{q_{\text{rev}}}{T_{\text{surr}}}$$

The magnitudes of $-q_{\text{rev}}$ and q_{rev} are equal, their opposite arithmetic signs denoting loss of heat by the system and gain of heat by the surroundings. Since $T_{\text{sys}} > T_{\text{surr}}$ in this scenario, the entropy *decrease* of the system will be less than the entropy *increase* of the surroundings, and so *the entropy of the universe will increase*:

Equation:

$$|\Delta S_{\text{sys}}| < |\Delta S_{\text{surr}}|$$
$$\Delta S_{\text{univ}} = \Delta S_{\text{sys}} + \Delta S_{\text{surr}} > 0$$

2. The objects are at different temperatures, and heat flows from the cooler to the hotter object. *This is never observed to occur spontaneously.* Again designating the hotter object as the system and invoking the definition of entropy yields the following:

Equation:

$$\Delta S_{\text{sys}} = \frac{q_{\text{rev}}}{T_{\text{sys}}} \quad \text{and} \quad \Delta S_{\text{surr}} = \frac{-q_{\text{rev}}}{T_{\text{surr}}}$$

The arithmetic signs of q_{rev} denote the gain of heat by the system and the loss of heat by the surroundings. The magnitude of the entropy change for the surroundings will again be greater than that for the system, but in this case, the signs of the heat changes (that is, *the direction of the heat flow*) will yield a negative value for ΔS_{univ} . *This process involves a decrease in the entropy of the universe.*

3. The objects are at essentially the same temperature, $T_{\text{sys}} \approx T_{\text{surr}}$, and so the magnitudes of the entropy changes are essentially the same for both the system and the surroundings. In this case, the entropy change of the universe is zero, and the system is *at equilibrium*.

Equation:

$$|\Delta S_{\text{sys}}| \approx |\Delta S_{\text{surr}}|$$

$$\Delta S_{\text{univ}} = \Delta S_{\text{sys}} + \Delta S_{\text{surr}} = 0$$

These results lead to a profound statement regarding the relation between entropy and spontaneity known as the **second law of thermodynamics**: *all spontaneous changes cause an increase in the entropy of the universe*. A summary of these three relations is provided in [\[link\]](#).

The Second Law of Thermodynamics	
$\Delta S_{\text{univ}} > 0$	spontaneous
$\Delta S_{\text{univ}} < 0$	nonspontaneous (spontaneous in opposite direction)
$\Delta S_{\text{univ}} = 0$	at equilibrium

For many realistic applications, the surroundings are vast in comparison to the system. In such cases, the heat gained or lost by the surroundings as a result of some process represents a very small, nearly infinitesimal, fraction of its total thermal energy. For example, combustion of a fuel in air involves transfer of heat from a system (the fuel and oxygen molecules undergoing reaction) to surroundings that are infinitely more massive (the earth's atmosphere). As a result, q_{surr} is a good approximation of q_{rev} , and the second law may be stated as the following:

Equation:

$$\Delta S_{\text{univ}} = \Delta S_{\text{sys}} + \Delta S_{\text{surr}} = \Delta S_{\text{sys}} + \frac{q_{\text{surr}}}{T}$$

We may use this equation to predict the spontaneity of a process as illustrated in [\[link\]](#).

Example:

Will Ice Spontaneously Melt?

The entropy change for the process

Equation:



is 22.1 J/K and requires that the surroundings transfer 6.00 kJ of heat to the system. Is the process spontaneous at -10.00°C ? Is it spontaneous at $+10.00^\circ\text{C}$?

Solution

We can assess the spontaneity of the process by calculating the entropy change of the universe. If ΔS_{univ} is positive, then the process is spontaneous. At both temperatures, $\Delta S_{\text{sys}} = 22.1 \text{ J/K}$ and $q_{\text{surr}} = -6.00 \text{ kJ}$. At -10.00°C (263.15 K), the following is true:

Equation:

$$\begin{aligned}\Delta S_{\text{univ}} &= \Delta S_{\text{sys}} + \Delta S_{\text{surr}} = \Delta S_{\text{sys}} + \frac{q_{\text{surr}}}{T} \\ &= 22.1 \text{ J/K} + \frac{-6.00 \times 10^3 \text{ J}}{263.15 \text{ K}} = -0.7 \text{ J/K}\end{aligned}$$

$S_{\text{univ}} < 0$, so melting is nonspontaneous (*not* spontaneous) at $-10.0\text{ }^{\circ}\text{C}$.

At $10.00\text{ }^{\circ}\text{C}$ (283.15 K), the following is true:

Equation:

$$\begin{aligned}\Delta S_{\text{univ}} &= \Delta S_{\text{sys}} + \frac{q_{\text{surr}}}{T} \\ &= 22.1\text{ J/K} + \frac{-6.00 \times 10^3\text{ J}}{283.15\text{ K}} = +0.9\text{ J/K}\end{aligned}$$

$S_{\text{univ}} > 0$, so melting is spontaneous at $10.00\text{ }^{\circ}\text{C}$.

Check Your Learning

Using this information, determine if liquid water will spontaneously freeze at the same temperatures. What can you say about the values of S_{univ} ?

Note:

Answer:

Entropy is a state function, so $\Delta S_{\text{freezing}} = -\Delta S_{\text{melting}} = -22.1\text{ J/K}$ and $q_{\text{surr}} = +6.00\text{ kJ}$. At $-10.00\text{ }^{\circ}\text{C}$ spontaneous, $+0.7\text{ J/K}$; at $+10.00\text{ }^{\circ}\text{C}$ nonspontaneous, -0.9 J/K .

The Third Law of Thermodynamics

The previous section described the various contributions of matter and energy dispersal that contribute to the entropy of a system. With these contributions in mind, consider the entropy of a pure, perfectly crystalline solid possessing no kinetic energy (that is, at a temperature of absolute zero, 0 K). This system may be described by a single microstate, as its purity, perfect crystallinity and complete lack of motion means there is but one possible location for each identical atom or molecule comprising the crystal ($W = 1$). According to the Boltzmann equation, the entropy of this system is zero.

Equation:

$$S = k \ln W = k \ln(1) = 0$$

This limiting condition for a system's entropy represents the **third law of thermodynamics**: *the entropy of a pure, perfect crystalline substance at 0 K is zero.*

Careful calorimetric measurements can be made to determine the temperature dependence of a substance's entropy and to derive absolute entropy values under specific conditions. **Standard entropies (S°)** are for one mole of substance under standard conditions (a pressure of 1 bar and a temperature of 298.15 K ; see details regarding standard conditions in the thermochemistry chapter of this text). The **standard entropy change (ΔS°)** for a reaction may be computed using standard entropies as shown below:

Equation:

$$\Delta S^{\circ} = \sum \nu S^{\circ}(\text{products}) - \sum \nu S^{\circ}(\text{reactants})$$

where ν represents stoichiometric coefficients in the balanced equation representing the process. For example, ΔS° for the following reaction at room temperature

Equation:



is computed as:

Equation:

$$= [xS^{\circ}(\text{C}) + yS^{\circ}(\text{D})] - [mS^{\circ}(\text{A}) + nS^{\circ}(\text{B})]$$

A partial listing of standard entropies is provided in [\[link\]](#), and additional values are provided in [Appendix G](#). The example exercises that follow demonstrate the use of S° values in calculating standard entropy changes for physical and chemical processes.

Substance	S° (J mol ⁻¹ K ⁻¹)
carbon	
C(s, graphite)	5.740
C(s, diamond)	2.38
CO(g)	197.7
CO ₂ (g)	213.8
CH ₄ (g)	186.3
C ₂ H ₄ (g)	219.5
C ₂ H ₆ (g)	229.5
CH ₃ OH(l)	126.8
C ₂ H ₅ OH(l)	160.7
hydrogen	
H ₂ (g)	130.57
H(g)	114.6
H ₂ O(g)	188.71
H ₂ O(l)	69.91
HCl(g)	186.8
H ₂ S(g)	205.7
oxygen	

O₂(g)

205.03

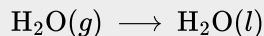
Standard entropies for selected substances measured at 1 atm and 298.15 K. (Values are approximately equal to those measured at 1 bar, the currently accepted standard state pressure.)

Example:

Determination of ΔS°

Calculate the standard entropy change for the following process:

Equation:



Solution

Calculate the entropy change using standard entropies as shown above:

Equation:

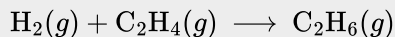
$$\Delta S^\circ = (1 \text{ mol})(70.0 \text{ J mol}^{-1} \text{ K}^{-1}) - (1 \text{ mol})(188.8 \text{ J mol}^{-1} \text{ K}^{-1}) = -118.8 \text{ J/K}$$

The value for ΔS° is negative, as expected for this phase transition (condensation), which the previous section discussed.

Check Your Learning

Calculate the standard entropy change for the following process:

Equation:



Note:

Answer:

$$-120.6 \text{ J K}^{-1} \text{ mol}^{-1}$$

Example:

Determination of ΔS°

Calculate the standard entropy change for the combustion of methanol, CH₃OH:

Equation:



Solution

Calculate the entropy change using standard entropies as shown above:

Equation:

$$\Delta S^\circ = \sum \nu S^\circ(\text{products}) - \sum \nu S^\circ(\text{reactants})$$

Equation:

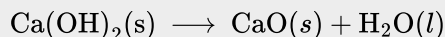
$$[2 \text{ mol} \times S^\circ(\text{CO}_2(g)) + 4 \text{ mol} \times S^\circ(\text{H}_2\text{O}(l))] - [2 \text{ mol} \times S^\circ(\text{CH}_3\text{OH}(l)) + 3 \text{ mol} \times S^\circ(\text{O}_2(g))]$$

$$= \{ [2(213.8) + 4 \times 70.0] - [2(126.8) + 3(205.03)] \} = -161.1 \text{ J/K}$$

Check Your Learning

Calculate the standard entropy change for the following reaction:

Equation:



Note:

Answer:

$$24.7 \text{ J K}^{-1} \text{ mol}^{-1}$$

Key Concepts and Summary

The second law of thermodynamics states that a spontaneous process increases the entropy of the universe, $S_{\text{univ}} > 0$. If $\Delta S_{\text{univ}} < 0$, the process is nonspontaneous, and if $\Delta S_{\text{univ}} = 0$, the system is at equilibrium. The third law of thermodynamics establishes the zero for entropy as that of a perfect, pure crystalline solid at 0 K. With only one possible microstate, the entropy is zero. We may compute the standard entropy change for a process by using standard entropy values for the reactants and products involved in the process.

Key Equations

- $\Delta S^\circ = \sum \nu S^\circ(\text{products}) - \sum \nu S^\circ(\text{reactants})$
- $\Delta S = \frac{q_{\text{rev}}}{T}$
- $\Delta S_{\text{univ}} = \Delta S_{\text{sys}} + \Delta S_{\text{surr}}$
- $\Delta S_{\text{univ}} = \Delta S_{\text{sys}} + \Delta S_{\text{surr}} = \Delta S_{\text{sys}} + \frac{q_{\text{surr}}}{T}$

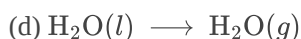
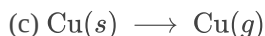
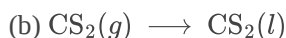
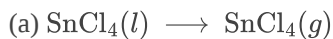
Chemistry End of Chapter Exercises

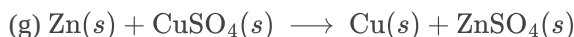
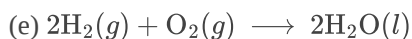
Exercise:

Problem: What is the difference between ΔS and ΔS° for a chemical change?

Exercise:

Problem: Calculate ΔS° for the following changes.





Solution:

(a) 107 J/K; (b) -86.4 J/K; (c) 133.2 J/K; (d) 118.8 J/K; (e) -326.6 J/K; (f) -171.9 J/K; (g) -7.2 J/K

Exercise:

Problem:

Determine the entropy change for the combustion of liquid ethanol, C_2H_5OH , under the standard conditions to give gaseous carbon dioxide and liquid water.

Exercise:

Problem:

Determine the entropy change for the combustion of gaseous propane, C_3H_8 , under the standard conditions to give gaseous carbon dioxide and water.

Solution:

100.6 J/K

Exercise:

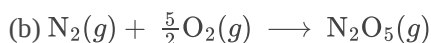
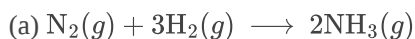
Problem:

“Thermite” reactions have been used for welding metal parts such as railway rails and in metal refining. One such thermite reaction is $Fe_2O_3(s) + 2Al(s) \longrightarrow Al_2O_3(s) + 2Fe(s)$. Is the reaction spontaneous at room temperature under standard conditions? During the reaction, the surroundings absorb 851.8 kJ/mol of heat.

Exercise:

Problem:

Using the relevant S° values listed in [Appendix G](#), calculate ΔS°_{298} for the following changes:

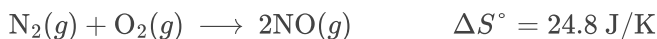
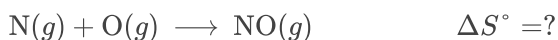


Solution:

(a) -198.1 J/K; (b) -348.9 J/K

Exercise:

Problem: From the following information, determine ΔS° for the following:





Exercise:

Problem:

By calculating ΔS_{univ} at each temperature, determine if the melting of 1 mole of NaCl(s) is spontaneous at 500 °C and at 700 °C.

$$S_{\text{NaCl}(s)}^\circ = 72.11 \frac{\text{J}}{\text{mol}\cdot\text{K}} \quad S_{\text{NaCl}(l)}^\circ = 95.06 \frac{\text{J}}{\text{mol}\cdot\text{K}} \quad \Delta H_{\text{fusion}}^\circ = 27.95 \text{ kJ/mol}$$

What assumptions are made about the thermodynamic information (entropy and enthalpy values) used to solve this problem?

Solution:

As $\Delta S_{\text{univ}} < 0$ at each of these temperatures, melting is not spontaneous at either of them. The given values for entropy and enthalpy are for NaCl at 298 K. It is assumed that these do not change significantly at the higher temperatures used in the problem.

Exercise:

Problem:

Use the standard entropy data in [Appendix G](#) to determine the change in entropy for each of the following reactions. All the processes occur at the standard conditions and 25 °C.

- (a) $\text{MnO}_2(s) \longrightarrow \text{Mn}(s) + \text{O}_2(g)$
- (b) $\text{H}_2(g) + \text{Br}_2(l) \longrightarrow 2\text{HBr}(g)$
- (c) $\text{Cu}(s) + \text{S}(g) \longrightarrow \text{CuS}(s)$
- (d) $2\text{LiOH}(s) + \text{CO}_2(g) \longrightarrow \text{Li}_2\text{CO}_3(s) + \text{H}_2\text{O}(g)$
- (e) $\text{CH}_4(g) + \text{O}_2(g) \longrightarrow \text{C}(s, \text{graphite}) + 2\text{H}_2\text{O}(g)$
- (f) $\text{CS}_2(g) + 3\text{Cl}_2(g) \longrightarrow \text{CCl}_4(g) + \text{S}_2\text{Cl}_2(g)$

Exercise:

Problem:

Use the standard entropy data in [Appendix G](#) to determine the change in entropy for each of the reactions listed in [\[link\]](#). All the processes occur at the standard conditions and 25 °C.

Solution:

- (a) 2.86 J/K; (b) 24.8 J/K; (c) -113.2 J/K; (d) -24.7 J/K; (e) 15.5 J/K; (f) 290.0 J/K

Glossary

second law of thermodynamics

all spontaneous processes involve an increase in the entropy of the universe

standard entropy (S°)

entropy for one mole of a substance at 1 bar pressure; tabulated values are usually determined at 298.15 K

standard entropy change (ΔS°)

change in entropy for a reaction calculated using the standard entropies

third law of thermodynamics

entropy of a perfect crystal at absolute zero (0 K) is zero

Free Energy

By the end of this section, you will be able to:

- Define Gibbs free energy, and describe its relation to spontaneity
- Calculate free energy change for a process using free energies of formation for its reactants and products
- Calculate free energy change for a process using enthalpies of formation and the entropies for its reactants and products
- Explain how temperature affects the spontaneity of some processes
- Relate standard free energy changes to equilibrium constants

One of the challenges of using the second law of thermodynamics to determine if a process is spontaneous is that it requires measurements of the entropy change for the system *and* the entropy change for the surroundings. An alternative approach involving a new thermodynamic property defined in terms of system properties only was introduced in the late nineteenth century by American mathematician Josiah Willard Gibbs. This new property is called the **Gibbs free energy (*G*)** (or simply the *free energy*), and it is defined in terms of a system's enthalpy and entropy as the following:

Equation:

$$G = H - TS$$

Free energy is a state function, and at constant temperature and pressure, the **free energy change (ΔG)** may be expressed as the following:

Equation:

$$\Delta G = \Delta H - T\Delta S$$

(For simplicity's sake, the subscript "sys" will be omitted henceforth.)

The relationship between this system property and the spontaneity of a process may be understood by recalling the previously derived second law expression:

Equation:

$$\Delta S_{\text{univ}} = \Delta S + \frac{q_{\text{surr}}}{T}$$

The first law requires that $q_{\text{surr}} = -q_{\text{sys}}$, and at constant pressure $q_{\text{sys}} = \Delta H$, so this expression may be rewritten as:

Equation:

$$\Delta S_{\text{univ}} = \Delta S - \frac{\Delta H}{T}$$

Multiplying both sides of this equation by $-T$, and rearranging yields the following:

Equation:

$$-T\Delta S_{\text{univ}} = \Delta H - T\Delta S$$

Comparing this equation to the previous one for free energy change shows the following relation:

Equation:

$$\Delta G = -T\Delta S_{\text{univ}}$$

The free energy change is therefore a reliable indicator of the spontaneity of a process, being directly related to the previously identified spontaneity indicator, ΔS_{univ} . [\[link\]](#) summarizes the relation between the spontaneity of a process and the arithmetic signs of these indicators.

Relation between Process Spontaneity and Signs of Thermodynamic Properties		
$\Delta S_{\text{univ}} > 0$	$\Delta G < 0$	spontaneous
$\Delta S_{\text{univ}} < 0$	$\Delta G > 0$	nonspontaneous
$\Delta S_{\text{univ}} = 0$	$\Delta G = 0$	at equilibrium

What's "Free" about ΔG ?

In addition to indicating spontaneity, the free energy change also provides information regarding the amount of useful work (w) that may be accomplished by a spontaneous process. Although a rigorous treatment of this subject is beyond the scope of an introductory chemistry text, a brief discussion is helpful for gaining a better perspective on this important thermodynamic property.

For this purpose, consider a spontaneous, exothermic process that involves a decrease in entropy. The free energy, as defined by

Equation:

$$\Delta G = \Delta H - T\Delta S$$

may be interpreted as representing the difference between the energy produced by the process, ΔH , and the energy lost to the surroundings, $T\Delta S$. The difference between the energy produced and the energy lost is the energy available (or "free") to do useful work by the process, ΔG . If the process somehow could be made to take place under conditions of thermodynamic reversibility, the amount of work that could be done would be maximal:

Equation:

$$\Delta G = w_{\text{max}}$$

However, as noted previously in this chapter, such conditions are not realistic. In addition, the technologies used to extract work from a spontaneous process (e.g., automobile engine, steam turbine) are never 100% efficient, and so the work done by these processes is always less than the theoretical maximum. Similar reasoning may be applied to a nonspontaneous process, for which the free energy change represents the *minimum* amount of work that must be done *on* the system to carry out the process.

Calculating Free Energy Change

Free energy is a state function, so its value depends only on the conditions of the initial and final states of the system. A convenient and common approach to the calculation of free energy changes for physical and chemical reactions is by use of widely available compilations of standard state thermodynamic data. One method involves the use of standard enthalpies and entropies to compute **standard free energy changes, ΔG°** , according to the following relation.

Equation:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

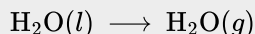
Example:

Using Standard Enthalpy and Entropy Changes to Calculate ΔG°

Use standard enthalpy and entropy data from [Appendix G](#) to calculate the standard free energy change for the vaporization of water at room temperature (298 K). What does the computed value for ΔG° say about the spontaneity of this process?

Solution

The process of interest is the following:

Equation:

The standard change in free energy may be calculated using the following equation:

Equation:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

From [Appendix G](#):

Substance	ΔH_f° (kJ/mol)	S° (J/K·mol)
$\text{H}_2\text{O}(l)$	-286.83	70.0
$\text{H}_2\text{O}(g)$	-241.82	188.8

Using the appendix data to calculate the standard enthalpy and entropy changes yields:

Equation:

$$\begin{aligned}\Delta H^\circ &= \Delta H_f^\circ (\text{H}_2\text{O}(g)) - \Delta H_f^\circ (\text{H}_2\text{O}(l)) \\ &= [-241.82 \text{ kJ/mol} - (-286.83)] \text{ kJ/mol} = 45.01 \text{ kJ}\end{aligned}$$

Equation:

$$\begin{aligned}\Delta S^\circ &= 1 \text{ mol} \times S^\circ (\text{H}_2\text{O}(g)) - 1 \text{ mol} \times S^\circ (\text{H}_2\text{O}(l)) \\ &= (1 \text{ mol})188.8 \text{ J/mol}\cdot\text{K} - (1 \text{ mol})70.0 \text{ J/mol}\cdot\text{K} = 118.8 \text{ J/K}\end{aligned}$$

Equation:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

Substitution into the standard free energy equation yields:

Equation:

$$\begin{aligned}\Delta G^\circ &= \Delta H^\circ - T\Delta S^\circ \\ &= 45.01 \text{ kJ} - (298 \text{ K} \times 118.8 \text{ J/K}) \times \frac{1 \text{ kJ}}{1000 \text{ J}}\end{aligned}$$

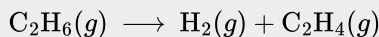
Equation:

$$45.01 \text{ kJ} - 35.4 \text{ kJ} = 9.6 \text{ kJ}$$

At 298 K (25 °C) $\Delta G^\circ > 0$, so boiling is nonspontaneous (*not* spontaneous).

Check Your Learning

Use standard enthalpy and entropy data from [Appendix G](#) to calculate the standard free energy change for the reaction shown here (298 K). What does the computed value for ΔG° say about the spontaneity of this process?

Equation:

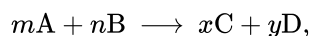
Note:

Answer:

$\Delta G^\circ = 102.0 \text{ kJ/mol}$; the reaction is nonspontaneous (*not* spontaneous) at 25 °C.

The standard free energy change for a reaction may also be calculated from **standard free energy of formation** ΔG_f° values of the reactants and products involved in the reaction. The standard free energy of formation is the free energy change that accompanies the formation of one mole of a substance from its elements in their standard states. Similar to the standard enthalpy of formation, ΔG_f° is by definition zero for elemental substances under standard state conditions. The approach used to calculate ΔG° for a reaction from ΔG_f° values is the same as that demonstrated previously for enthalpy and entropy changes. For the reaction

Equation:



the standard free energy change at room temperature may be calculated as

Equation:

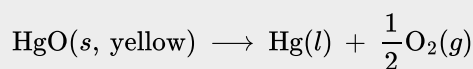
$$\begin{aligned}\Delta G^\circ &= \sum v\Delta G^\circ(\text{products}) - \sum v\Delta G^\circ(\text{reactants}) \\ &= [x\Delta G_f^\circ(\text{C}) + y\Delta G_f^\circ(\text{D})] - [m\Delta G_f^\circ(\text{A}) + n\Delta G_f^\circ(\text{B})].\end{aligned}$$

Example:

Using Standard Free Energies of Formation to Calculate ΔG°

Consider the decomposition of yellow mercury(II) oxide.

Equation:



Calculate the standard free energy change at room temperature, ΔG° , using (a) standard free energies of formation and (b) standard enthalpies of formation and standard entropies. Do the results indicate the reaction to be spontaneous or nonspontaneous under standard conditions?

Solution

The required data are available in [Appendix G](#) and are shown here.

Compound	ΔG_f° (kJ/mol)	ΔH_f° (kJ/mol)	S° (J/K·mol)
HgO (s, yellow)	-58.43	-90.46	71.13

Compound	ΔG_f° (kJ/mol)	ΔH_f° (kJ/mol)	S° (J/K·mol)
Hg(<i>l</i>)	0	0	75.9
O ₂ (<i>g</i>)	0	0	205.2

(a) Using free energies of formation:

Equation:

$$\Delta G^\circ = \sum \nu G_f^\circ(\text{products}) - \sum \nu \Delta G_f^\circ(\text{reactants})$$

Equation:

$$= \left[1 \Delta G_f^\circ \text{Hg}(\textit{l}) + \frac{1}{2} \Delta G_f^\circ \text{O}_2(\textit{g}) \right] - 1 \Delta G_f^\circ \text{HgO}(\textit{s}, \text{yellow})$$

Equation:

$$= \left[1 \text{ mol}(0 \text{ kJ/mol}) + \frac{1}{2} \text{ mol}(0 \text{ kJ/mol}) \right] - 1 \text{ mol}(-58.43 \text{ kJ/mol}) = 58.43 \text{ kJ/mol}$$

(b) Using enthalpies and entropies of formation:

Equation:

$$\Delta H^\circ = \sum \nu \Delta H_f^\circ(\text{products}) - \sum \nu \Delta H_f^\circ(\text{reactants})$$

Equation:

$$= \left[1 \Delta H_f^\circ \text{Hg}(\textit{l}) + \frac{1}{2} \Delta H_f^\circ \text{O}_2(\textit{g}) \right] - 1 \Delta H_f^\circ \text{HgO}(\textit{s}, \text{yellow})$$

Equation:

$$= \left[1 \text{ mol}(0 \text{ kJ/mol}) + \frac{1}{2} \text{ mol}(0 \text{ kJ/mol}) \right] - 1 \text{ mol}(-90.46 \text{ kJ/mol}) = 90.46 \text{ kJ/mol}$$

Equation:

$$\Delta S^\circ = \sum \nu \Delta S^\circ(\text{products}) - \sum \nu \Delta S^\circ(\text{reactants})$$

Equation:

$$= \left[1 \Delta S^\circ \text{Hg}(\textit{l}) + \frac{1}{2} \Delta S^\circ \text{O}_2(\textit{g}) \right] - 1 \Delta S^\circ \text{HgO}(\textit{s}, \text{yellow})$$

Equation:

$$= \left[1 \text{ mol}(75.9 \text{ J/mol K}) + \frac{1}{2} \text{ mol}(205.2 \text{ J/mol K}) \right] - 1 \text{ mol}(71.13 \text{ J/mol K}) = 107.4 \text{ J/mol K}$$

Equation:

$$\Delta G^\circ = \Delta H^\circ - T \Delta S^\circ = 90.46 \text{ kJ} - 298.15 \text{ K} \times 107.4 \text{ J/K} \cdot \text{mol} \times \frac{1 \text{ kJ}}{1000 \text{ J}}$$

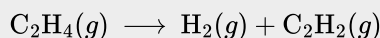
Equation:

$$\Delta G^\circ = (90.46 - 32.01) \text{ kJ/mol} = 58.45 \text{ kJ/mol}$$

Both ways to calculate the standard free energy change at 25 °C give the same numerical value (to three significant figures), and both predict that the process is nonspontaneous (*not* spontaneous) at room temperature.

Check Your Learning

Calculate ΔG° using (a) free energies of formation and (b) enthalpies of formation and entropies ([Appendix G](#)). Do the results indicate the reaction to be spontaneous or nonspontaneous at 25 °C?

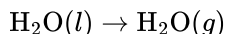
Equation:**Note:****Answer:**

(a) 140.8 kJ/mol, nonspontaneous

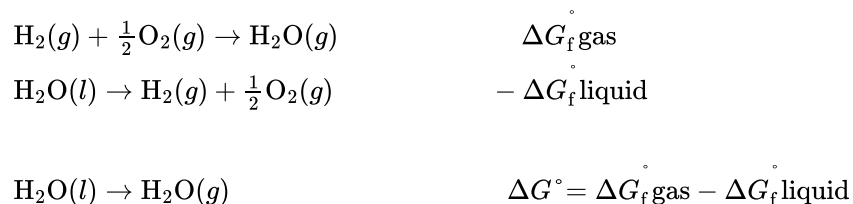
(b) 141.5 kJ/mol, nonspontaneous

Free Energy Changes for Coupled Reactions

The use of free energies of formation to compute free energy changes for reactions as described above is possible because ΔG is a state function, and the approach is analogous to the use of Hess' Law in computing enthalpy changes (see the chapter on thermochemistry). Consider the vaporization of water as an example:

Equation:

An equation representing this process may be derived by adding the formation reactions for the two phases of water (necessarily reversing the reaction for the liquid phase). The free energy change for the sum reaction is the sum of free energy changes for the two added reactions:

Equation:

This approach may also be used in cases where a nonspontaneous reaction is enabled by coupling it to a spontaneous reaction. For example, the production of elemental zinc from zinc sulfide is thermodynamically unfavorable, as indicated by a positive value for ΔG° :

Equation:

The industrial process for production of zinc from sulfidic ores involves coupling this decomposition reaction to the thermodynamically favorable oxidation of sulfur:

Equation:



The coupled reaction exhibits a negative free energy change and is spontaneous:

Equation:



This process is typically carried out at elevated temperatures, so this result obtained using standard free energy values is just an estimate. The gist of the calculation, however, holds true.

Example:

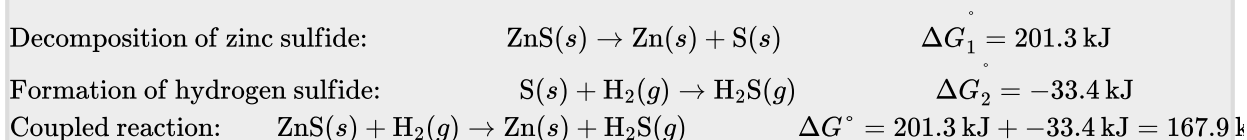
Calculating Free Energy Change for a Coupled Reaction

Is a reaction coupling the decomposition of ZnS to the formation of H₂S expected to be spontaneous under standard conditions?

Solution

Following the approach outlined above and using free energy values from [Appendix G](#):

Equation:

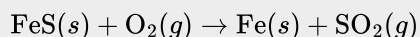


The coupled reaction exhibits a positive free energy change and is thus nonspontaneous.

Check Your Learning

What is the standard free energy change for the reaction below? Is the reaction expected to be spontaneous under standard conditions?

Equation:



Note:

Answer:

-199.7 kJ; spontaneous

Temperature Dependence of Spontaneity

As was previously demonstrated in this chapter's section on entropy, the spontaneity of a process may depend upon the temperature of the system. Phase transitions, for example, will proceed spontaneously in one direction or the other depending upon the temperature of the substance in question. Likewise, some chemical reactions can also

exhibit temperature dependent spontaneities. To illustrate this concept, the equation relating free energy change to the enthalpy and entropy changes for the process is considered:

Equation:

$$\Delta G = \Delta H - T\Delta S$$

The spontaneity of a process, as reflected in the arithmetic sign of its free energy change, is then determined by the signs of the enthalpy and entropy changes and, in some cases, the absolute temperature. Since T is the absolute (kelvin) temperature, it can only have positive values. Four possibilities therefore exist with regard to the signs of the enthalpy and entropy changes:

1. **Both ΔH and ΔS are positive.** This condition describes an endothermic process that involves an increase in system entropy. In this case, ΔG will be negative if the magnitude of the $T\Delta S$ term is greater than ΔH . If the $T\Delta S$ term is less than ΔH , the free energy change will be positive. Such a process is *spontaneous at high temperatures and nonspontaneous at low temperatures*.
2. **Both ΔH and ΔS are negative.** This condition describes an exothermic process that involves a decrease in system entropy. In this case, ΔG will be negative if the magnitude of the $T\Delta S$ term is less than ΔH . If the $T\Delta S$ term's magnitude is greater than ΔH , the free energy change will be positive. Such a process is *spontaneous at low temperatures and nonspontaneous at high temperatures*.
3. **ΔH is positive and ΔS is negative.** This condition describes an endothermic process that involves a decrease in system entropy. In this case, ΔG will be positive regardless of the temperature. Such a process is *nonspontaneous at all temperatures*.
4. **ΔH is negative and ΔS is positive.** This condition describes an exothermic process that involves an increase in system entropy. In this case, ΔG will be negative regardless of the temperature. Such a process is *spontaneous at all temperatures*.

These four scenarios are summarized in [\[link\]](#).

Summary of the Four Scenarios for Enthalpy and Entropy Changes

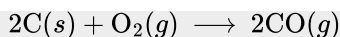
	$\Delta H > 0$ (endothermic)	$\Delta H < 0$ (exothermic)
$\Delta S > 0$ (increase in entropy)	$\Delta G < 0$ at high temperature $\Delta G > 0$ at low temperature Process is spontaneous at high temperature	$\Delta G < 0$ at any temperature Process is spontaneous at any temperature
$\Delta S < 0$ (decrease in entropy)	$\Delta G > 0$ at any temperature Process is nonspontaneous at any temperature	$\Delta G < 0$ at low temperature $\Delta G > 0$ at high temperature Process is spontaneous at low temperature

There are four possibilities regarding the signs of enthalpy and entropy changes.

Example:
Predicting the Temperature Dependence of Spontaneity

The incomplete combustion of carbon is described by the following equation:

Equation:



How does the spontaneity of this process depend upon temperature?

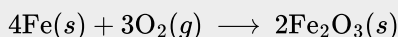
Solution

Combustion processes are exothermic ($\Delta H < 0$). This particular reaction involves an increase in entropy due to the accompanying increase in the amount of gaseous species (net gain of one mole of gas, $\Delta S > 0$). The reaction is therefore spontaneous ($\Delta G < 0$) at all temperatures.

Check Your Learning

Popular chemical hand warmers generate heat by the air-oxidation of iron:

Equation:



How does the spontaneity of this process depend upon temperature?

Note:

Answer:

ΔH and ΔS are negative; the reaction is spontaneous at low temperatures.

When considering the conclusions drawn regarding the temperature dependence of spontaneity, it is important to keep in mind what the terms “high” and “low” mean. Since these terms are adjectives, the temperatures in question are deemed high or low relative to some reference temperature. A process that is nonspontaneous at one temperature but spontaneous at another will necessarily undergo a change in “spontaneity” (as reflected by its ΔG) as temperature varies. This is clearly illustrated by a graphical presentation of the free energy change equation, in which ΔG is plotted on the y axis versus T on the x axis:

Equation:

$$\Delta G = \Delta H - T\Delta S$$

Equation:

$$y = b + mx$$

Such a plot is shown in [\[link\]](#). A process whose enthalpy and entropy changes are of the same arithmetic sign will exhibit a temperature-dependent spontaneity as depicted by the two yellow lines in the plot. Each line crosses from one spontaneity domain (positive or negative ΔG) to the other at a temperature that is characteristic of the process in question. This temperature is represented by the x-intercept of the line, that is, the value of T for which ΔG is zero:

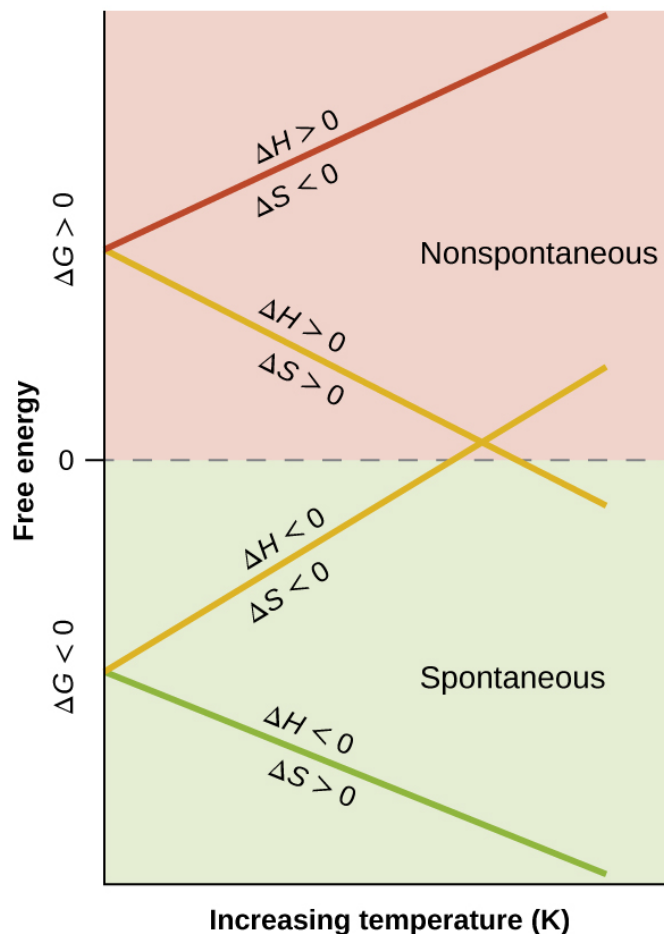
Equation:

$$\Delta G = 0 = \Delta H - T\Delta S$$

Equation:

$$T = \frac{\Delta H}{\Delta S}$$

So, saying a process is spontaneous at “high” or “low” temperatures means the temperature is above or below, respectively, that temperature at which ΔG for the process is zero. As noted earlier, the condition of $\Delta G = 0$ describes a system at equilibrium.



These plots show the variation in ΔG with temperature for the four possible combinations of arithmetic sign for ΔH and ΔS .

Example:

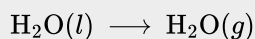
Equilibrium Temperature for a Phase Transition

As defined in the chapter on liquids and solids, the boiling point of a liquid is the temperature at which its liquid and gaseous phases are in equilibrium (that is, when vaporization and condensation occur at equal rates). Use the information in [Appendix G](#) to estimate the boiling point of water.

Solution

The process of interest is the following phase change:

Equation:



When this process is at equilibrium, $\Delta G = 0$, so the following is true:

Equation:

$$0 = \Delta H^\circ - T\Delta S^\circ \quad \text{or} \quad T = \frac{\Delta H^\circ}{\Delta S^\circ}$$

Using the standard thermodynamic data from [Appendix G](#),

Equation:

$$\begin{aligned}\Delta H^\circ &= 1 \text{ mol} \times \Delta H_f^\circ (\text{H}_2\text{O}(g)) - 1 \text{ mol} \times \Delta H_f^\circ (\text{H}_2\text{O}(l)) \\ &= (1 \text{ mol}) - 241.82 \text{ kJ/mol} - (1 \text{ mol}) (-286.83 \text{ kJ/mol}) = 44.01 \text{ kJ}\end{aligned}$$

Equation:

$$\begin{aligned}\Delta S^\circ &= 1 \text{ mol} \times \Delta S^\circ (\text{H}_2\text{O}(g)) - 1 \text{ mol} \times \Delta S^\circ (\text{H}_2\text{O}(l)) \\ &= (1 \text{ mol}) 188.8 \text{ J/K}\cdot\text{mol} - (1 \text{ mol}) 70.0 \text{ J/K}\cdot\text{mol} = 118.8 \text{ J/K}\end{aligned}$$

Equation:

$$T = \frac{\Delta H^\circ}{\Delta S^\circ} = \frac{44.01 \times 10^3 \text{ J}}{118.8 \text{ J/K}} = 370.5 \text{ K} = 97.3^\circ \text{C}$$

The accepted value for water's normal boiling point is 373.2 K (100.0 °C), and so this calculation is in reasonable agreement. Note that the values for enthalpy and entropy changes data used were derived from standard data at 298 K ([Appendix G](#)). If desired, you could obtain more accurate results by using enthalpy and entropy changes determined at (or at least closer to) the actual boiling point.

Check Your Learning

Use the information in [Appendix G](#) to estimate the boiling point of CS₂.

Note:

Answer:

313 K (accepted value 319 K)

Free Energy and Equilibrium

The free energy change for a process may be viewed as a measure of its driving force. A negative value for ΔG represents a driving force for the process in the forward direction, while a positive value represents a driving force for the process in the reverse direction. When ΔG is zero, the forward and reverse driving forces are equal, and the process occurs in both directions at the same rate (the system is at equilibrium).

In the chapter on equilibrium the *reaction quotient*, Q , was introduced as a convenient measure of the status of an equilibrium system. Recall that Q is the numerical value of the mass action expression for the system, and that you may use its value to identify the direction in which a reaction will proceed in order to achieve equilibrium. When Q is lesser than the equilibrium constant, K , the reaction will proceed in the forward direction until equilibrium is reached and $Q = K$. Conversely, if $Q > K$, the process will proceed in the reverse direction until equilibrium is achieved.

The free energy change for a process taking place with reactants and products present under *nonstandard conditions* (pressures other than 1 bar; concentrations other than 1 M) is related to the standard free energy change according to this equation:

Equation:

$$\Delta G = \Delta G^\circ + RT \ln Q$$

R is the gas constant (8.314 J/K mol), T is the kelvin or absolute temperature, and Q is the reaction quotient. This equation may be used to predict the spontaneity for a process under any given set of conditions as illustrated in [\[link\]](#).

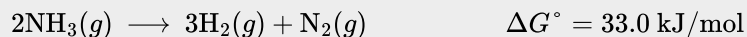
Example:

Calculating ΔG under Nonstandard Conditions

What is the free energy change for the process shown here under the specified conditions?

$T = 25\text{ }^{\circ}\text{C}$, $P_{\text{N}_2} = 0.870\text{ atm}$, $P_{\text{H}_2} = 0.250\text{ atm}$, and $P_{\text{NH}_3} = 12.9\text{ atm}$

Equation:



Solution

The equation relating free energy change to standard free energy change and reaction quotient may be used directly:

Equation:

$$\begin{aligned}\Delta G &= \Delta G^{\circ} + RT \ln Q = 33.0 \frac{\text{kJ}}{\text{mol}} + \left(8.314 \frac{\text{J}}{\text{mol K}} \times 298\text{ K} \times \ln \frac{(0.250^3) \times 0.870}{12.9^2} \right) \\ &= 9680 \frac{\text{J}}{\text{mol}} \text{ or } 9.68\text{ kJ/mol}\end{aligned}$$

Since the computed value for ΔG is positive, the reaction is nonspontaneous under these conditions.

Check Your Learning

Calculate the free energy change for this same reaction at $875\text{ }^{\circ}\text{C}$ in a 5.00 L mixture containing 0.100 mol of each gas. Is the reaction spontaneous under these conditions?

Note:

Answer:

$\Delta G = +45.1\text{ kJ/mol}$; no

For a system at equilibrium, $Q = K$ and $\Delta G = 0$, and the previous equation may be written as

Equation:

$$0 = \Delta G^{\circ} + RT \ln K \qquad (\text{at equilibrium})$$

Equation:

$$\Delta G^{\circ} = -RT \ln K \qquad \text{or} \qquad K = e^{-\frac{\Delta G^{\circ}}{RT}}$$

This form of the equation provides a useful link between these two essential thermodynamic properties, and it can be used to derive equilibrium constants from standard free energy changes and vice versa. The relations between standard free energy changes and equilibrium constants are summarized in [\[link\]](#).

Relations between Standard Free Energy Changes and Equilibrium Constants

K	ΔG°	Composition of an Equilibrium Mixture
> 1	< 0	Products are more abundant
< 1	> 0	Reactants are more abundant
$= 1$	$= 0$	Reactants and products are comparably abundant

Example:

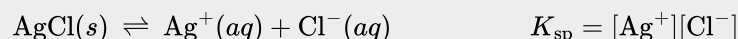
Calculating an Equilibrium Constant using Standard Free Energy Change

Given that the standard free energies of formation of $\text{Ag}^+(aq)$, $\text{Cl}^-(aq)$, and $\text{AgCl}(s)$ are 77.1 kJ/mol, -131.2 kJ/mol, and -109.8 kJ/mol, respectively, calculate the solubility product, K_{sp} , for AgCl .

Solution

The reaction of interest is the following:

Equation:



The standard free energy change for this reaction is first computed using standard free energies of formation for its reactants and products:

Equation:

$$\begin{aligned} \Delta G^\circ &= [\Delta G_f^\circ (\text{Ag}^+(aq)) + \Delta G_f^\circ (\text{Cl}^-(aq))] - [\Delta G_f^\circ (\text{AgCl}(s))] \\ &= [77.1 \text{ kJ/mol} - 131.2 \text{ kJ/mol}] - [-109.8 \text{ kJ/mol}] = 55.7 \text{ kJ/mol} \end{aligned}$$

The equilibrium constant for the reaction may then be derived from its standard free energy change:

Equation:

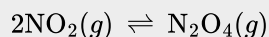
$$K_{\text{sp}} = e^{-\frac{\Delta G^\circ}{RT}} = \exp\left(-\frac{\Delta G^\circ}{RT}\right) = \exp\left(-\frac{55.7 \times 10^3 \text{ J/mol}}{8.314 \text{ J/mol}\cdot\text{K} \times 298.15 \text{ K}}\right) = \exp(-22.470) = e^{-22.470} = 1.$$

This result is in reasonable agreement with the value provided in [Appendix J](#).

Check Your Learning

Use the thermodynamic data provided in [Appendix G](#) to calculate the equilibrium constant for the dissociation of dinitrogen tetroxide at 25 °C.

Equation:

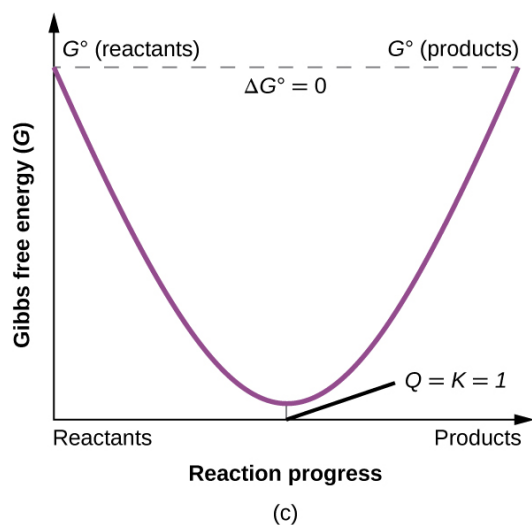
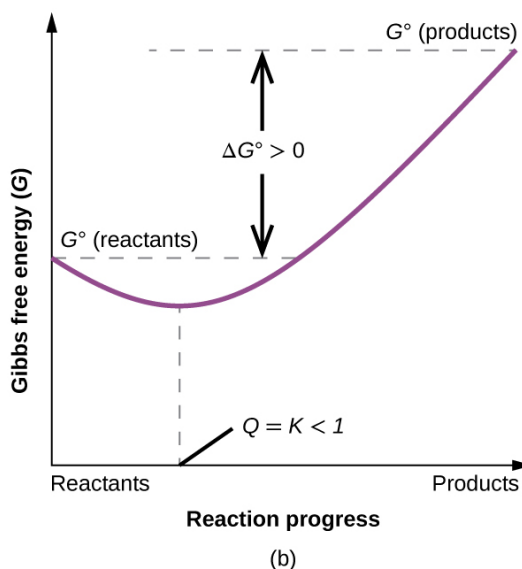
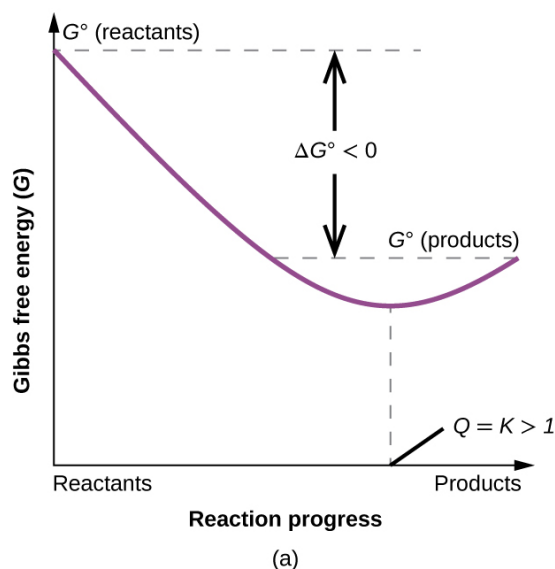


Note:

Answer:

$$K = 6.9$$

To further illustrate the relation between these two essential thermodynamic concepts, consider the observation that reactions spontaneously proceed in a direction that ultimately establishes equilibrium. As may be shown by plotting the free energy change versus the extent of the reaction (for example, as reflected in the value of Q), equilibrium is established when the system's free energy is minimized ([link](#)). If a system consists of reactants and products in nonequilibrium amounts ($Q \neq K$), the reaction will proceed spontaneously in the direction necessary to establish equilibrium.



These plots show the free energy versus reaction progress for systems whose standard free energy changes are (a) negative, (b) positive, and (c) zero. Nonequilibrium systems will proceed spontaneously in whatever direction is necessary to minimize free energy and establish equilibrium.

Key Concepts and Summary

Gibbs free energy (G) is a state function defined with regard to system quantities only and may be used to predict the spontaneity of a process. A negative value for ΔG indicates a spontaneous process; a positive ΔG indicates a nonspontaneous process; and a ΔG of zero indicates that the system is at equilibrium. A number of approaches to the computation of free energy changes are possible.

Key Equations

- $\Delta G = \Delta H - T\Delta S$

Chemistry End of Chapter Exercises

Exercise:

Problem: What is the difference between ΔG and ΔG° for a chemical change?

Exercise:

Problem:

A reaction has $\Delta H^\circ = 100 \text{ kJ/mol}$ and $\Delta S^\circ = 250 \text{ J/mol}\cdot\text{K}$. Is the reaction spontaneous at room temperature? If not, under what temperature conditions will it become spontaneous?

Solution:

The reaction is nonspontaneous at room temperature.
Above 400 K, ΔG will become negative, and the reaction will become spontaneous.

Exercise:

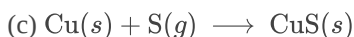
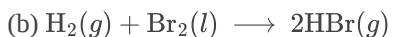
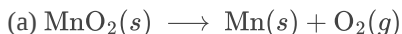
Problem:

Explain what happens as a reaction starts with $\Delta G < 0$ (negative) and reaches the point where $\Delta G = 0$.

Exercise:

Problem:

Use the standard free energy of formation data in [Appendix G](#) to determine the free energy change for each of the following reactions, which are run under standard state conditions and 25 °C. Identify each as either spontaneous or nonspontaneous at these conditions.



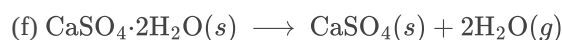
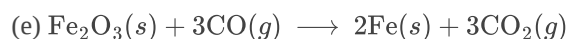
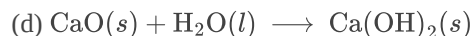
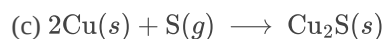
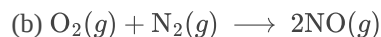
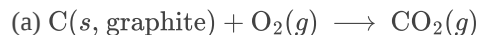
Solution:

(a) 465.1 kJ nonspontaneous; (b) -106.86 kJ spontaneous; (c) -291.9 kJ spontaneous; (d) -83.4 kJ spontaneous; (e) -406.7 kJ spontaneous; (f) -30.0 kJ spontaneous

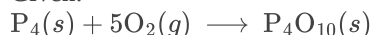
Exercise:

Problem:

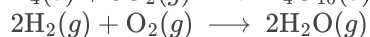
Use the standard free energy data in [Appendix G](#) to determine the free energy change for each of the following reactions, which are run under standard state conditions and 25 °C. Identify each as either spontaneous or nonspontaneous at these conditions.

**Exercise:**

Given:



$$\Delta G^\circ = -2697.0 \text{ kJ/mol}$$



$$\Delta G^\circ = -457.18 \text{ kJ/mol}$$



$$\Delta G^\circ = -428.66 \text{ kJ/mol}$$

(a) Determine the standard free energy of formation, ΔG_f° , for phosphoric acid.

(b) How does your calculated result compare to the value in [Appendix G](#)? Explain.

Solution:

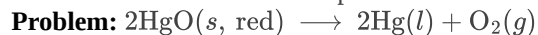
(a) The standard free energy of formation is -1124.3 kJ/mol . (b) The calculation agrees with the value in [Appendix G](#) because free energy is a state function (just like the enthalpy and entropy), so its change depends only on the initial and final states, not the path between them.

Exercise:**Problem:**

Is the formation of ozone ($\text{O}_3(g)$) from oxygen ($\text{O}_2(g)$) spontaneous at room temperature under standard state conditions?

Exercise:

Consider the decomposition of red mercury(II) oxide under standard state conditions.



(a) Is the decomposition spontaneous under standard state conditions?

(b) Above what temperature does the reaction become spontaneous?

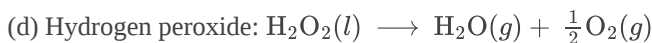
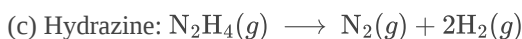
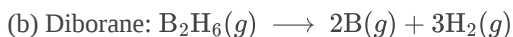
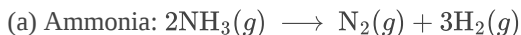
Solution:

(a) The reaction is nonspontaneous; (b) Above 566 °C the process is spontaneous.

Exercise:

Problem:

Among other things, an ideal fuel for the control thrusters of a space vehicle should decompose in a spontaneous exothermic reaction when exposed to the appropriate catalyst. Evaluate the following substances under standard state conditions as suitable candidates for fuels.

**Exercise:****Problem:**

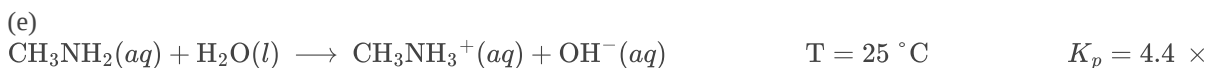
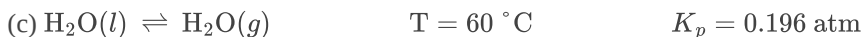
Calculate ΔG° for each of the following reactions from the equilibrium constant at the temperature given.

**Solution:**

(a) 1.5×10^2 kJ; (b) -21.9 kJ; (c) -5.34 kJ; (d) -0.383 kJ; (e) 18 kJ; (f) 71 kJ

Exercise:**Problem:**

Calculate ΔG° for each of the following reactions from the equilibrium constant at the temperature given.

**Exercise:****Problem:**

Calculate the equilibrium constant at 25°C for each of the following reactions from the value of ΔG° given.

- (a) $\text{O}_2(g) + 2\text{F}_2(g) \longrightarrow 2\text{OF}_2(g) \quad \Delta G^\circ = -9.2 \text{ kJ}$
- (b) $\text{I}_2(s) + \text{Br}_2(l) \longrightarrow 2\text{IBr}(g) \quad \Delta G^\circ = 7.3 \text{ kJ}$
- (c) $2\text{LiOH}(s) + \text{CO}_2(g) \longrightarrow \text{Li}_2\text{CO}_3(s) + \text{H}_2\text{O}(g) \quad \Delta G^\circ = -79 \text{ kJ}$
- (d) $\text{N}_2\text{O}_3(g) \longrightarrow \text{NO}(g) + \text{NO}_2(g) \quad \Delta G^\circ = -1.6 \text{ kJ}$
- (e) $\text{SnCl}_4(l) \longrightarrow \text{SnCl}_4(g) \quad \Delta G^\circ = 8.0 \text{ kJ}$
-

Solution:

- (a) $K = 41$; (b) $K = 0.053$; (c) $K = 6.9 \times 10^{13}$; (d) $K = 1.9$; (e) $K = 0.04$

Exercise:

Problem:

Calculate the equilibrium constant at 25 °C for each of the following reactions from the value of ΔG° given.

- (a) $\text{I}_2(s) + \text{Cl}_2(g) \longrightarrow 2\text{ICl}(g) \quad \Delta G^\circ = -10.88 \text{ kJ}$
- (b) $\text{H}_2(g) + \text{I}_2(s) \longrightarrow 2\text{HI}(g) \quad \Delta G^\circ = 3.4 \text{ kJ}$
- (c) $\text{CS}_2(g) + 3\text{Cl}_2(g) \longrightarrow \text{CCl}_4(g) + \text{S}_2\text{Cl}_2(g) \quad \Delta G^\circ = -39 \text{ kJ}$
- (d) $2\text{SO}_2(g) + \text{O}_2(g) \longrightarrow 2\text{SO}_3(g) \quad \Delta G^\circ = -141.82 \text{ kJ}$
- (e) $\text{CS}_2(g) \longrightarrow \text{CS}_2(l) \quad \Delta G^\circ = -1.88 \text{ kJ}$

Exercise:

Problem: Calculate the equilibrium constant at the temperature given.

- (a) $\text{O}_2(g) + 2\text{F}_2(g) \longrightarrow 2\text{F}_2\text{O}(g) \quad (\text{T} = 100^\circ \text{C})$
- (b) $\text{I}_2(s) + \text{Br}_2(l) \longrightarrow 2\text{IBr}(g) \quad (\text{T} = 0.0^\circ \text{C})$
- (c) $2\text{LiOH}(s) + \text{CO}_2(g) \longrightarrow \text{Li}_2\text{CO}_3(s) + \text{H}_2\text{O}(g) \quad (\text{T} = 575^\circ \text{C})$
- (d) $\text{N}_2\text{O}_3(g) \longrightarrow \text{NO}(g) + \text{NO}_2(g) \quad (\text{T} = -10.0^\circ \text{C})$
- (e) $\text{SnCl}_4(l) \longrightarrow \text{SnCl}_4(g) \quad (\text{T} = 200^\circ \text{C})$
-

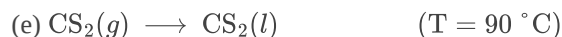
Solution:

In each of the following, the value of ΔG is not given at the temperature of the reaction. Therefore, we must calculate ΔG from the values ΔH° and ΔS and then calculate ΔG from the relation $\Delta G = \Delta H^\circ - T\Delta S^\circ$. (a) $K = 1.29$; (b) $K = 2.51 \times 10^{-3}$; (c) $K = 4.83 \times 10^3$; (d) $K = 0.219$; (e) $K = 16.1$

Exercise:

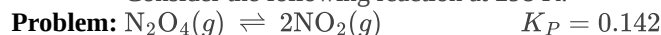
Problem: Calculate the equilibrium constant at the temperature given.

- (a) $\text{I}_2(s) + \text{Cl}_2(g) \longrightarrow 2\text{ICl}(g) \quad (\text{T} = 100^\circ \text{C})$
- (b) $\text{H}_2(g) + \text{I}_2(s) \longrightarrow 2\text{HI}(g) \quad (\text{T} = 0.0^\circ \text{C})$
- (c) $\text{CS}_2(g) + 3\text{Cl}_2(g) \longrightarrow \text{CCl}_4(g) + \text{S}_2\text{Cl}_2(g) \quad (\text{T} = 125^\circ \text{C})$



Exercise:

Consider the following reaction at 298 K:



What is the standard free energy change at this temperature? Describe what happens to the initial system, where the reactants and products are in standard states, as it approaches equilibrium.

Solution:

The standard free energy change is $\Delta G^\circ = -RT \ln K = 4.84 \text{ kJ/mol}$. When reactants and products are in their standard states (1 bar or 1 atm), $Q = 1$. As the reaction proceeds toward equilibrium, the reaction shifts left (the amount of products drops while the amount of reactants increases): $Q < 1$, and ΔG becomes less positive as it approaches zero. At equilibrium, $Q = K$, and $\Delta G = 0$.

Exercise:

Problem:

Determine the normal boiling point (in kelvin) of dichloroethane, CH_2Cl_2 . Find the actual boiling point using the Internet or some other source, and calculate the percent error in the temperature. Explain the differences, if any, between the two values.

Exercise:

Problem: Under what conditions is $\text{N}_2\text{O}_3(g) \longrightarrow \text{NO}(g) + \text{NO}_2(g)$ spontaneous?

Solution:

The reaction will be spontaneous at temperatures greater than 287 K.

Exercise:

Problem:

At room temperature, the equilibrium constant (K_w) for the self-ionization of water is 1.00×10^{-14} . Using this information, calculate the standard free energy change for the aqueous reaction of hydrogen ion with hydroxide ion to produce water. (Hint: The reaction is the reverse of the self-ionization reaction.)

Exercise:

Problem:

Hydrogen sulfide is a pollutant found in natural gas. Following its removal, it is converted to sulfur by the reaction $2\text{H}_2\text{S}(g) + \text{SO}_2(g) \rightleftharpoons \frac{3}{8}\text{S}_8(s, \text{ rhombic}) + 2\text{H}_2\text{O}(l)$. What is the equilibrium constant for this reaction? Is the reaction endothermic or exothermic?

Solution:

$K = 5.35 \times 10^{15}$; The process is exothermic.

Exercise:

Problem:

Consider the decomposition of $\text{CaCO}_3(s)$ into $\text{CaO}(s)$ and $\text{CO}_2(g)$. What is the equilibrium partial pressure of CO_2 at room temperature?

Exercise:**Problem:**

In the laboratory, hydrogen chloride ($\text{HCl}(g)$) and ammonia ($\text{NH}_3(g)$) often escape from bottles of their solutions and react to form the ammonium chloride ($\text{NH}_4\text{Cl}(s)$), the white glaze often seen on glassware. Assuming that the number of moles of each gas that escapes into the room is the same, what is the maximum partial pressure of HCl and NH_3 in the laboratory at room temperature? (Hint: The partial pressures will be equal and are at their maximum value when at equilibrium.)

Solution:

1.0×10^{-8} atm. This is the maximum pressure of the gases under the stated conditions.

Exercise:**Problem:**

Benzene can be prepared from acetylene. $3\text{C}_2\text{H}_2(g) \rightleftharpoons \text{C}_6\text{H}_6(g)$. Determine the equilibrium constant at 25 °C and at 850 °C. Is the reaction spontaneous at either of these temperatures? Why is all acetylene not found as benzene?

Exercise:**Problem:**

Carbon dioxide decomposes into CO and O_2 at elevated temperatures. What is the equilibrium partial pressure of oxygen in a sample at 1000 °C for which the initial pressure of CO_2 was 1.15 atm?

Solution:

$$x = 1.29 \times 10^{-5} \text{ atm} = P_{\text{O}_2}$$

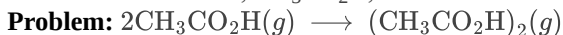
Exercise:**Problem:**

Carbon tetrachloride, an important industrial solvent, is prepared by the chlorination of methane at 850 K. $\text{CH}_4(g) + 4\text{Cl}_2(g) \longrightarrow \text{CCl}_4(g) + 4\text{HCl}(g)$

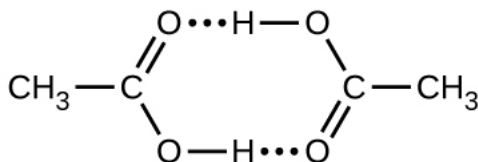
What is the equilibrium constant for the reaction at 850 K? Would the reaction vessel need to be heated or cooled to keep the temperature of the reaction constant?

Exercise:

Acetic acid, $\text{CH}_3\text{CO}_2\text{H}$, can form a dimer, $(\text{CH}_3\text{CO}_2\text{H})_2$, in the gas phase.



The dimer is held together by two hydrogen bonds with a total strength of 66.5 kJ per mole of dimer.



At 25 °C, the equilibrium constant for the dimerization is 1.3×10^3 (pressure in atm). What is ΔS° for the reaction?

Solution:

-0.16 kJ

Exercise:

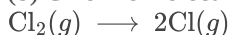
Problem: Determine ΔG° for the following reactions.

(a) Antimony pentachloride decomposes at 448 °C. The reaction is:



An equilibrium mixture in a 5.00 L flask at 448 °C contains 3.85 g of SbCl_5 , 9.14 g of SbCl_3 , and 2.84 g of Cl_2 .

(b) Chlorine molecules dissociate according to this reaction:



1.00% of Cl_2 molecules dissociate at 975 K and a pressure of 1.00 atm.

Solution:

(a) 22.1 kJ; (b) 61.5 kJ/mol

Exercise:

Problem:

Given that the ΔG_f° for $\text{Pb}^{2+}(aq)$ and $\text{Cl}^-(aq)$ is -24.3 kJ/mole and -131.2 kJ/mole respectively, determine the solubility product, K_{sp} , for $\text{PbCl}_2(s)$.

Exercise:

Problem:

Determine the standard free energy change, ΔG_f° , for the formation of $\text{S}^{2-}(aq)$ given that the ΔG_f° for $\text{Ag}^+(aq)$ and $\text{Ag}_2\text{S}(s)$ are 77.1 kJ/mole and -39.5 kJ/mole respectively, and the solubility product for $\text{Ag}_2\text{S}(s)$ is 8×10^{-51} .

Solution:

90 kJ/mol

Exercise:

Problem:

Determine the standard enthalpy change, entropy change, and free energy change for the conversion of diamond to graphite. Discuss the spontaneity of the conversion with respect to the enthalpy and entropy changes. Explain why diamond spontaneously changing into graphite is not observed.

Exercise:

The evaporation of one mole of water at 298 K has a standard free energy change of 8.58 kJ.

Problem: $\text{H}_2\text{O}(l) \rightleftharpoons \text{H}_2\text{O}(g)$ $\Delta G^\circ = 8.58 \text{ kJ}$

(a) Is the evaporation of water under standard thermodynamic conditions spontaneous?

(b) Determine the equilibrium constant, K_P , for this physical process.

(c) By calculating ΔG , determine if the evaporation of water at 298 K is spontaneous when the partial pressure of water, $P_{\text{H}_2\text{O}}$, is 0.011 atm.

(d) If the evaporation of water were always nonspontaneous at room temperature, wet laundry would never dry when placed outside. In order for laundry to dry, what must be the value of $P_{\text{H}_2\text{O}}$ in the air?

Solution:

(a) Under standard thermodynamic conditions, the evaporation is nonspontaneous; (b) $K_p = 0.031$; (c) The evaporation of water is spontaneous; (d) $P_{\text{H}_2\text{O}}$ must always be less than K_p or less than 0.031 atm. 0.031 atm represents air saturated with water vapor at 25 °C, or 100% humidity.

Exercise:

Problem:

In glycolysis, the reaction of glucose (Glu) to form glucose-6-phosphate (G6P) requires ATP to be present as described by the following equation:



In this process, ATP becomes ADP summarized by the following equation:



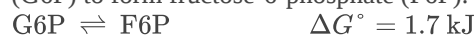
Determine the standard free energy change for the following reaction, and explain why ATP is necessary to drive this process:



Exercise:

Problem:

One of the important reactions in the biochemical pathway glycolysis is the reaction of glucose-6-phosphate (G6P) to form fructose-6-phosphate (F6P):



(a) Is the reaction spontaneous or nonspontaneous under standard thermodynamic conditions?

(b) Standard thermodynamic conditions imply the concentrations of G6P and F6P to be 1 M, however, in a typical cell, they are not even close to these values. Calculate ΔG when the concentrations of G6P and F6P are 120 μM and 28 μM respectively, and discuss the spontaneity of the forward reaction under these conditions. Assume the temperature is 37 °C.

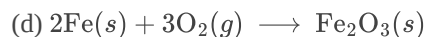
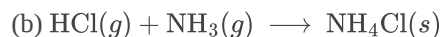
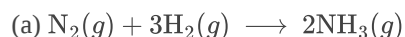
Solution:

(a) Nonspontaneous as $\Delta G^\circ > 0$; (b) $\Delta G = \Delta G^\circ + RT \ln Q$,
 $\Delta G = 1.7 \times 10^3 + (8.314 \times 310 \times \ln \frac{28}{120}) = -2.1 \text{ kJ}$. The forward reaction to produce F6P is spontaneous under these conditions.

Exercise:

Problem:

Without doing a numerical calculation, determine which of the following will reduce the free energy change for the reaction, that is, make it less positive or more negative, when the temperature is increased. Explain.



Exercise:

Problem:

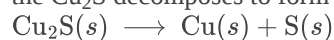
When ammonium chloride is added to water and stirred, it dissolves spontaneously and the resulting solution feels cold. Without doing any calculations, deduce the signs of ΔG , ΔH , and ΔS for this process, and justify your choices.

Solution:

ΔG is negative as the process is spontaneous. ΔH is positive as with the solution becoming cold, the dissolving must be endothermic. ΔS must be positive as this drives the process, and it is expected for the dissolution of any soluble ionic compound.

Exercise:**Problem:**

An important source of copper is from the copper ore, chalcocite, a form of copper(I) sulfide. When heated, the Cu_2S decomposes to form copper and sulfur described by the following equation:



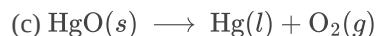
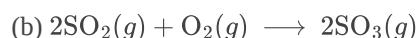
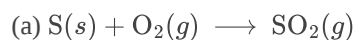
(a) Determine ΔG° for the decomposition of $\text{Cu}_2\text{S}(s)$.

(b) The reaction of sulfur with oxygen yields sulfur dioxide as the only product. Write an equation that describes this reaction, and determine ΔG° for the process.

(c) The production of copper from chalcocite is performed by roasting the Cu_2S in air to produce the Cu. By combining the equations from Parts (a) and (b), write the equation that describes the roasting of the chalcocite, and explain why coupling these reactions together makes for a more efficient process for the production of the copper.

Exercise:**Problem:**

What happens to ΔG (becomes more negative or more positive) for the following chemical reactions when the partial pressure of oxygen is increased?

**Solution:**

(a) Increasing the oxygen partial pressure will yield a decrease in Q and ΔG thus becomes more negative. (b) Increasing the oxygen partial pressure will yield a decrease in Q and ΔG thus becomes more negative. (c) Increasing the oxygen partial pressure will yield an increase in Q and ΔG thus becomes more positive.

Glossary**Gibbs free energy change (G)**

thermodynamic property defined in terms of system enthalpy and entropy; all spontaneous processes involve a decrease in G

standard free energy change (ΔG°)

change in free energy for a process occurring under standard conditions (1 bar pressure for gases, 1 M concentration for solutions)

standard free energy of formation (ΔG_f°)
change in free energy accompanying the formation of one mole of substance from its elements in their
standard states

Electrode and Cell Potentials

By the end of this section, you will be able to:

- Describe and relate the definitions of electrode and cell potentials
- Interpret electrode potentials in terms of relative oxidant and reductant strengths
- Calculate cell potentials and predict redox spontaneity using standard electrode potentials

Unlike the spontaneous oxidation of copper by aqueous silver(I) ions described in section 17.2, immersing a copper wire in an aqueous solution of lead(II) ions yields no reaction. The two species, $\text{Ag}^+(aq)$ and $\text{Pb}^{2+}(aq)$, thus show a distinct difference in their redox activity towards copper: the silver ion spontaneously oxidized copper, but the lead ion did not. Electrochemical cells permit this relative redox activity to be quantified by an easily measured property, *potential*. This property is more commonly called *voltage* when referenced in regard to electrical applications, and it is a measure of energy accompanying the transfer of charge. Potentials are measured in the volt unit, defined as one joule of energy per one coulomb of charge, $V = J/C$.

When measured for purposes of electrochemistry, a potential reflects the driving force for a specific type of charge transfer process, namely, the transfer of electrons between redox reactants. Considering the nature of potential in this context, it is clear that the potential of a single half-cell or a single electrode can't be measured; "transfer" of electrons requires both a donor and recipient, in this case a reductant and an oxidant, respectively. Instead, a half-cell potential may only be assessed relative to that of another half-cell. It is only the *difference in potential* between two half-cells that may be measured, and these measured potentials are called **cell potentials**, E_{cell} , defined as

Equation:

$$E_{\text{cell}} = E_{\text{cathode}} - E_{\text{anode}}$$

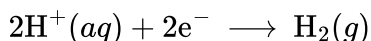
where E_{cathode} and E_{anode} are the potentials of two different half-cells functioning as specified in the subscripts. As for other thermodynamic quantities, the **standard cell potential**, E°_{cell} , is a cell potential measured when both half-cells are under standard-state conditions (1 M concentrations, 1 bar pressures, 298 K):

Equation:

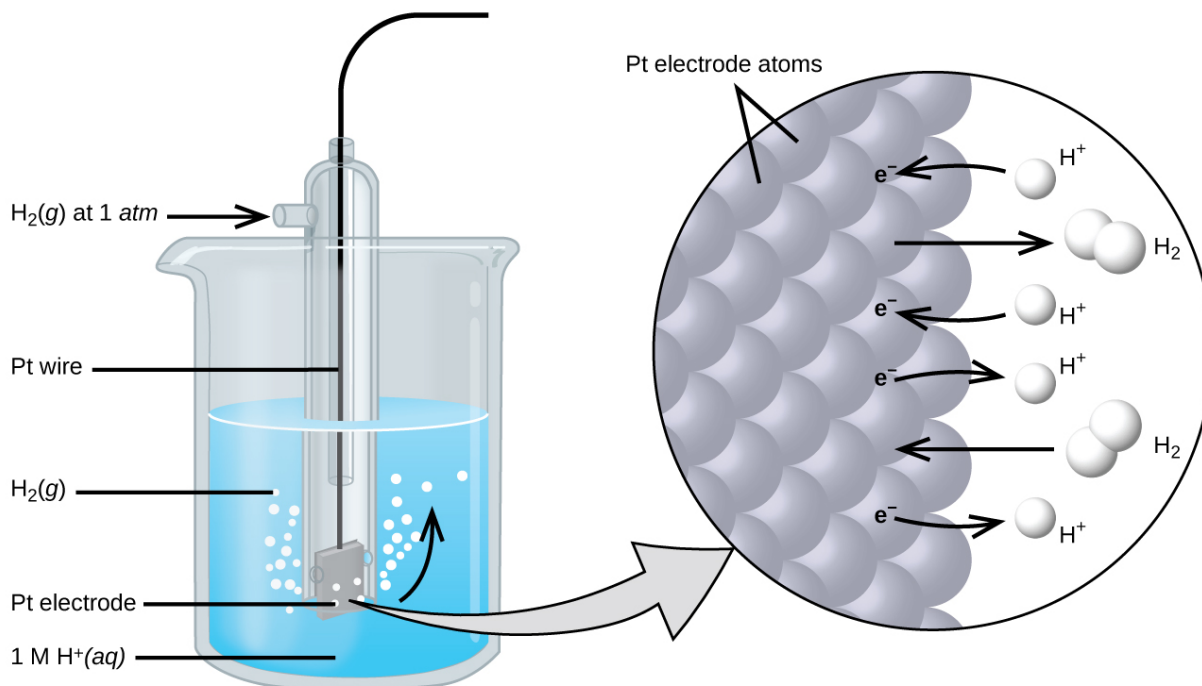
$$E^{\circ}_{\text{cell}} = E^{\circ}_{\text{cathode}} - E^{\circ}_{\text{anode}}$$

To simplify the collection and sharing of potential data for half-reactions, the scientific community has designated one particular half-cell to serve as a universal reference for cell potential measurements, assigning it a potential of exactly 0 V. This half-cell is the **standard hydrogen electrode (SHE)** and it is based on half-reaction below:

Equation:



A typical SHE contains an inert platinum electrode immersed in precisely 1 M aqueous H^+ and a stream of bubbling H_2 gas at 1 bar pressure, all maintained at a temperature of 298 K (see [\[link\]](#)).



A standard hydrogen electrode (SHE).

The assigned potential of the SHE permits the definition of a conveniently measured potential for a single half-cell. The **electrode potential (E_X)** for a half-cell X is defined as *the potential measured for a cell comprised of X acting as cathode and the SHE acting as anode*:

Equation:

$$E_{\text{cell}} = E_X - E_{\text{SHE}}$$

$$E_{\text{SHE}} = 0 \text{ V (defined)}$$

$$E_{\text{cell}} = E_X$$

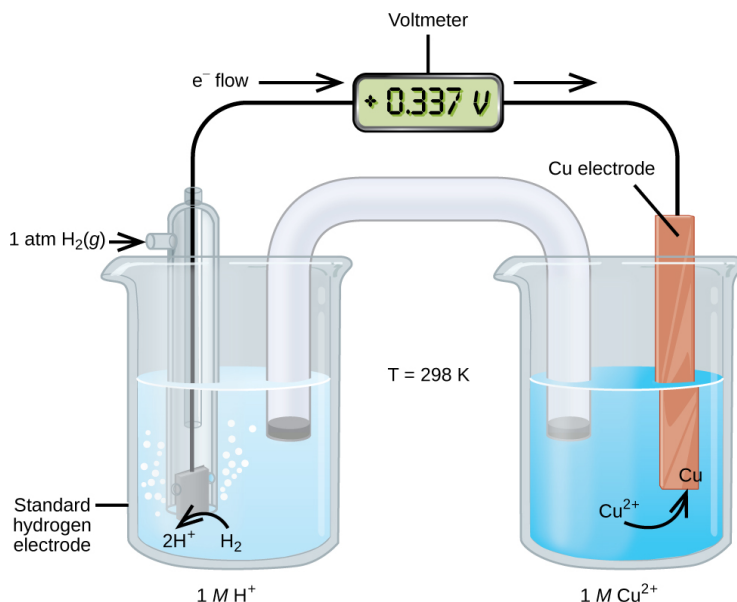
When the half-cell X is under standard-state conditions, its potential is the **standard electrode potential, E°_X** . Since the definition of cell potential requires the half-cells function as cathodes, these potentials are sometimes called *standard reduction potentials*.

This approach to measuring electrode potentials is illustrated in [\[link\]](#), which depicts a cell comprised of an SHE connected to a copper(II)/copper(0) half-cell under standard-state conditions. A voltmeter in the external circuit allows measurement of the potential difference between the two half-cells. Since the Cu half-cell is designated as the cathode in the definition of cell potential, it is connected to the red (positive) input of the voltmeter, while the designated SHE anode is connected to the black (negative) input. These connections insure that the sign of the measured potential will be consistent with the sign conventions of electrochemistry per the various definitions discussed above. A cell potential of +0.337 V is measured, and so

Equation:

$$E^{\circ}_{\text{cell}} = E^{\circ}_{\text{Cu}} = +0.337 \text{ V}$$

Tabulations of E° values for other half-cells measured in a similar fashion are available as reference literature to permit calculations of cell potentials and the prediction of the spontaneity of redox processes.



A cell permitting experimental measurement of the standard electrode potential for the half-reaction
 $\text{Cu}^{2+}(\text{aq}) + 2\text{e}^{-} \longrightarrow \text{Cu}(\text{s})$

[\[link\]](#) provides a listing of standard electrode potentials for a selection of half-reactions in numerical order, and a more extensive alphabetical listing is given in [Appendix L](#).

Selected Standard Reduction Potentials at 25 °C	
Half-Reaction	E° (V)
$\text{F}_2(\text{g}) + 2\text{e}^{-} \longrightarrow 2\text{F}^{-}(\text{aq})$	+2.866
$\text{PbO}_2(\text{s}) + \text{SO}_4^{2-}(\text{aq}) + 4\text{H}^{+}(\text{aq}) + 2\text{e}^{-} \longrightarrow \text{PbSO}_4(\text{s}) + 2\text{H}_2\text{O}(\text{l})$	+1.69
$\text{MnO}_4^{-}(\text{aq}) + 8\text{H}^{+}(\text{aq}) + 5\text{e}^{-} \longrightarrow \text{Mn}^{2+}(\text{aq}) + 4\text{H}_2\text{O}(\text{l})$	+1.507

Selected Standard Reduction Potentials at 25 °C	
Half-Reaction	E° (V)
$\text{Au}^{3+}(aq) + 3e^- \longrightarrow \text{Au}(s)$	+1.498
$\text{Cl}_2(g) + 2e^- \longrightarrow 2\text{Cl}^-(aq)$	+1.35827
$\text{O}_2(g) + 4\text{H}^+(aq) + 4e^- \longrightarrow 2\text{H}_2\text{O}(l)$	+1.229
$\text{Pt}^{2+}(aq) + 2e^- \longrightarrow \text{Pt}(s)$	+1.20
$\text{Br}_2(aq) + 2e^- \longrightarrow 2\text{Br}^-(aq)$	+1.0873
$\text{Ag}^+(aq) + e^- \longrightarrow \text{Ag}(s)$	+0.7996
$\text{Hg}_2^{2+}(aq) + 2e^- \longrightarrow 2\text{Hg}(l)$	+0.7973
$\text{Fe}^{3+}(aq) + e^- \longrightarrow \text{Fe}^{2+}(aq)$	+0.771
$\text{MnO}_4^-(aq) + 2\text{H}_2\text{O}(l) + 3e^- \longrightarrow \text{MnO}_2(s) + 4\text{OH}^-(aq)$	+0.558
$\text{I}_2(s) + 2e^- \longrightarrow 2\text{I}^-(aq)$	+0.5355
$\text{NiO}_2(s) + 2\text{H}_2\text{O}(l) + 2e^- \longrightarrow \text{Ni}(\text{OH})_2(s) + 2\text{OH}^-(aq)$	+0.49
$\text{Cu}^{2+}(aq) + 2e^- \longrightarrow \text{Cu}(s)$	+0.34
$\text{Hg}_2\text{Cl}_2(s) + 2e^- \longrightarrow 2\text{Hg}(l) + 2\text{Cl}^-(aq)$	+0.26808
$\text{AgCl}(s) + e^- \longrightarrow \text{Ag}(s) + \text{Cl}^-(aq)$	+0.22233
$\text{Sn}^{4+}(aq) + 2e^- \longrightarrow \text{Sn}^{2+}(aq)$	+0.151
$2\text{H}^+(aq) + 2e^- \longrightarrow \text{H}_2(g)$	0.00
$\text{Pb}^{2+}(aq) + 2e^- \longrightarrow \text{Pb}(s)$	-0.1262
$\text{Sn}^{2+}(aq) + 2e^- \longrightarrow \text{Sn}(s)$	-0.1375
$\text{Ni}^{2+}(aq) + 2e^- \longrightarrow \text{Ni}(s)$	-0.257
$\text{Co}^{2+}(aq) + 2e^- \longrightarrow \text{Co}(s)$	-0.28
$\text{PbSO}_4(s) + 2e^- \longrightarrow \text{Pb}(s) + \text{SO}_4^{2-}(aq)$	-0.3505

Selected Standard Reduction Potentials at 25 °C	
Half-Reaction	E° (V)
$\text{Cd}^{2+}(\text{aq}) + 2\text{e}^- \longrightarrow \text{Cd}(\text{s})$	-0.4030
$\text{Fe}^{2+}(\text{aq}) + 2\text{e}^- \longrightarrow \text{Fe}(\text{s})$	-0.447
$\text{Cr}^{3+}(\text{aq}) + 3\text{e}^- \longrightarrow \text{Cr}(\text{s})$	-0.744
$\text{Mn}^{2+}(\text{aq}) + 2\text{e}^- \longrightarrow \text{Mn}(\text{s})$	-1.185
$\text{Zn}(\text{OH})_2(\text{s}) + 2\text{e}^- \longrightarrow \text{Zn}(\text{s}) + 2\text{OH}^-(\text{aq})$	-1.245
$\text{Zn}^{2+}(\text{aq}) + 2\text{e}^- \longrightarrow \text{Zn}(\text{s})$	-0.7618
$\text{Al}^{3+}(\text{aq}) + 3\text{e}^- \longrightarrow \text{Al}(\text{s})$	-1.662
$\text{Mg}^{2+}(\text{aq}) + 2\text{e}^- \longrightarrow \text{Mg}(\text{s})$	-2.372
$\text{Na}^+(\text{aq}) + \text{e}^- \longrightarrow \text{Na}(\text{s})$	-2.71
$\text{Ca}^{2+}(\text{aq}) + 2\text{e}^- \longrightarrow \text{Ca}(\text{s})$	-2.868
$\text{Ba}^{2+}(\text{aq}) + 2\text{e}^- \longrightarrow \text{Ba}(\text{s})$	-2.912
$\text{K}^+(\text{aq}) + \text{e}^- \longrightarrow \text{K}(\text{s})$	-2.931
$\text{Li}^+(\text{aq}) + \text{e}^- \longrightarrow \text{Li}(\text{s})$	-3.04

Example:

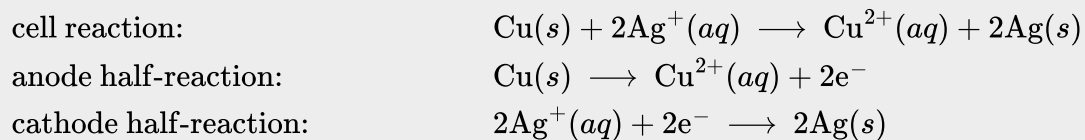
Calculating Standard Cell Potentials

What is the standard potential of the galvanic cell shown in [\[link\]](#)?

Solution

The cell in [\[link\]](#) is galvanic, the spontaneous cell reaction involving oxidation of its copper anode and reduction of silver(I) ions at its silver cathode:

Equation:



The standard cell potential computed as

Equation:

$$\begin{aligned}
 E^{\circ}_{\text{cell}} &= E^{\circ}_{\text{cathode}} - E^{\circ}_{\text{anode}} \\
 &= E^{\circ}_{\text{Ag}} - E^{\circ}_{\text{Cu}} \\
 &= 0.7996 \text{ V} - 0.34 \text{ V} \\
 &= +0.46 \text{ V}
 \end{aligned}$$

Check Your Learning

What is the standard cell potential expected if the silver cathode half-cell in [\[link\]](#) is replaced with a lead half-cell: $\text{Pb}^{2+}(\text{aq}) + 2\text{e}^{-} \longrightarrow \text{Pb}(\text{s})$?

Note:

Answer:

−0.47 V

Interpreting Electrode and Cell Potentials

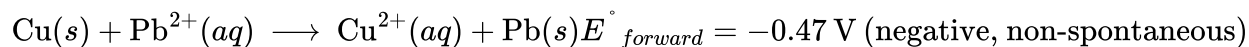
Thinking carefully about the definitions of cell and electrode potentials and the observations of spontaneous redox change presented thus far, a significant relation is noted. The previous section described the spontaneous oxidation of copper by aqueous silver(I) ions, but no observed reaction with aqueous lead(II) ions. Results of the calculations in [\[link\]](#) have just shown *the spontaneous process is described by a positive cell potential while the nonspontaneous process exhibits a negative cell potential*. And so, with regard to the relative effectiveness (“strength”) with which aqueous Ag^{+} and Pb^{2+} ions oxidize Cu under standard conditions, *the stronger oxidant is the one exhibiting the greater standard electrode potential, E°* . Since by convention electrode potentials are for reduction processes, an increased value of E° corresponds to an increased driving force behind the reduction of the species (hence increased effectiveness of its action as an *oxidizing agent* on some other species). Negative values for electrode potentials are simply a consequence of assigning a value of 0 V to the SHE, indicating the reactant of the half-reaction is a weaker oxidant than aqueous hydrogen ions.

Applying this logic to the numerically ordered listing of standard electrode potentials in [\[link\]](#) shows this listing to be likewise in order of the oxidizing strength of the half-reaction’s reactant species, decreasing from strongest oxidant (most positive E°) to weakest oxidant (most negative E°). Predictions regarding the spontaneity of redox reactions under standard state conditions can then be easily made by simply comparing the relative positions of their table entries. By definition, E°_{cell} is positive when $E^{\circ}_{\text{cathode}} > E^{\circ}_{\text{anode}}$, and so any redox reaction in which the oxidant’s entry is above the reductant’s entry is predicted to be spontaneous.

Reconsideration of the two redox reactions in [\[link\]](#) provides support for this fact. The entry for the silver(I)/silver(0) half-reaction is above that for the copper(II)/copper(0) half-reaction, and so the oxidation of Cu by Ag^{+} is predicted to be spontaneous ($E^{\circ}_{\text{cathode}} > E^{\circ}_{\text{anode}}$ and so $E^{\circ}_{\text{cell}} > 0$). Conversely, the entry for the lead(II)/lead(0) half-cell is beneath that for copper(II)/copper(0), and the oxidation of Cu by Pb^{2+} is nonspontaneous ($E^{\circ}_{\text{cathode}} < E^{\circ}_{\text{anode}}$ and so $E^{\circ}_{\text{cell}} < 0$).

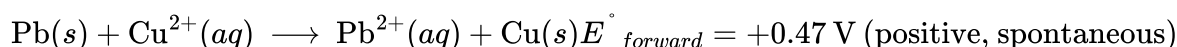
Recalling the chapter on thermodynamics, the spontaneities of the forward and reverse reactions of a reversible process show a reciprocal relationship: if a process is spontaneous in one direction, it is non-spontaneous in the opposite direction. As an indicator of spontaneity for redox reactions, the potential of a cell reaction shows a consequential relationship in its arithmetic sign. The spontaneous oxidation of copper by lead(II) ions is *not* observed,

Equation:



and so the reverse reaction, the oxidation of lead by copper(II) ions, is predicted to occur spontaneously:

Equation:



Note that reversing the direction of a redox reaction effectively interchanges the identities of the cathode and anode half-reactions, and so the cell potential is calculated from electrode potentials in the reverse subtraction order than that for the forward reaction. In practice, a voltmeter would report a potential of -0.47 V with its red and black inputs connected to the Pb and Cu electrodes, respectively. If the inputs were swapped, the reported voltage would be $+0.47 \text{ V}$.

Example:

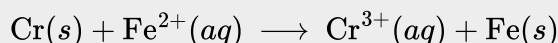
Predicting Redox Spontaneity

Are aqueous iron(II) ions predicted to spontaneously oxidize elemental chromium under standard state conditions? Assume the half-reactions to be those available in [\[link\]](#).

Solution

Referring to the tabulated half-reactions, the redox reaction in question can be represented by the equations below:

Equation:



The entry for the putative oxidant, Fe^{2+} , appears *above* the entry for the reductant, Cr, and so a spontaneous reaction is predicted per the quick approach described above. Supporting this predication by calculating the standard cell potential for this reaction gives

Equation:

$$\begin{aligned} E^\circ_{\text{cell}} &= E^\circ_{\text{cathode}} - E^\circ_{\text{anode}} \\ &= E^\circ_{\text{Fe(II)}} - E^\circ_{\text{Cr}} \\ &= -0.447 \text{ V} - -0.774 \text{ V} = +0.297 \text{ V} \end{aligned}$$

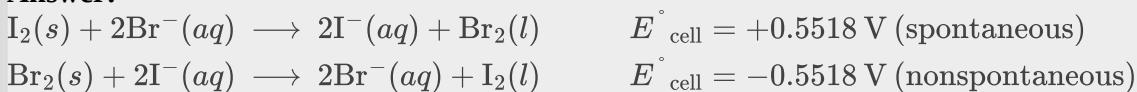
The positive value for the standard cell potential indicates the process is spontaneous under standard state conditions.

Check Your Learning

Use the data in [\[link\]](#) to predict the spontaneity of the oxidation of bromide ion by molecular iodine under standard state conditions, supporting the prediction by calculating the standard cell potential for the reaction. Repeat for the oxidation of iodide ion by molecular bromine.

Note:

Answer:



Key Concepts and Summary

The property of potential, E , is the energy associated with the separation/transfer of charge. In electrochemistry, the potentials of cells and half-cells are thermodynamic quantities that reflect the driving force or the spontaneity of their redox processes. The cell potential of an electrochemical cell is the difference in between its cathode and anode. To permit easy sharing of half-cell potential data, the standard hydrogen electrode (SHE) is assigned a potential of exactly 0 V and used to define a single electrode potential for any given half-cell. The electrode potential of a half-cell, E_X , is the cell potential of said half-cell acting as a cathode when connected to a SHE acting as an anode. When the half-cell is operating under standard state conditions, its potential is the standard electrode potential, E°_X . Standard electrode potentials reflect the relative oxidizing strength of the half-reaction's reactant, with stronger oxidants exhibiting larger (more positive) E°_X values. Tabulations of standard electrode potentials may be used to compute standard cell potentials, E°_{cell} , for many redox reactions. The arithmetic sign of a cell potential indicates the spontaneity of the cell reaction, with positive values for spontaneous reactions and negative values for nonspontaneous reactions (spontaneous in the reverse direction).

Key Equations

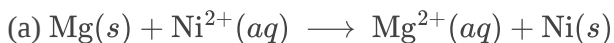
$$\bullet E^\circ_{\text{cell}} = E^\circ_{\text{cathode}} - E^\circ_{\text{anode}}$$

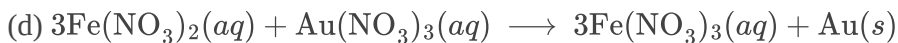
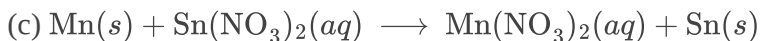
Chemistry End of Chapter Exercises

Exercise:

Problem:

Calculate the standard cell potential for each reaction below, and note whether the reaction is spontaneous under standard state conditions.





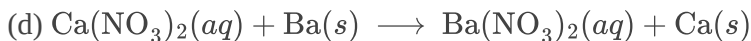
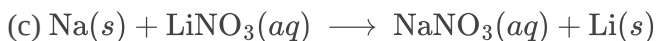
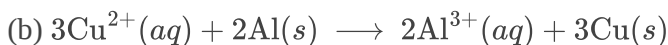
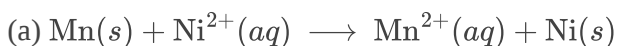
Solution:

(a) +2.115 V (spontaneous); (b) +0.4626 V (spontaneous); (c) +1.0589 V (spontaneous); (d) +0.727 V (spontaneous)

Exercise:

Problem:

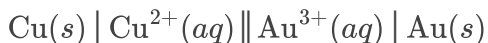
Calculate the standard cell potential for each reaction below, and note whether the reaction is spontaneous under standard state conditions.



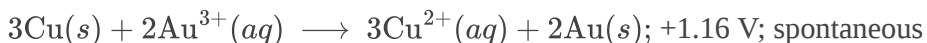
Exercise:

Problem:

Write the balanced cell reaction for the cell schematic below, calculate the standard cell potential, and note whether the reaction is spontaneous under standard state conditions.



Solution:



Exercise:

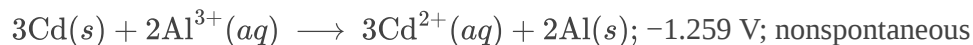
Problem:

Determine the cell reaction and standard cell potential at 25 °C for a cell made from a cathode half-cell consisting of a silver electrode in 1 M silver nitrate solution and an anode half-cell consisting of a zinc electrode in 1 M zinc nitrate. Is the reaction spontaneous at standard conditions?

Exercise:

Problem:

Determine the cell reaction and standard cell potential at 25 °C for a cell made from an anode half-cell containing a cadmium electrode in 1 M cadmium nitrate and an anode half-cell consisting of an aluminum electrode in 1 M aluminum nitrate solution. Is the reaction spontaneous at standard conditions?

Solution:**Exercise:****Problem:**

Write the balanced cell reaction for the cell schematic below, calculate the standard cell potential, and note whether the reaction is spontaneous under standard state conditions.

**Glossary**

standard cell potential (E_{cell}°)

the cell potential when all reactants and products are in their standard states (1 bar or 1 atm or gases; 1 *M* for solutes), usually at 298.15 K

standard hydrogen electrode (SHE)

half-cell based on hydrogen ion production, assigned a potential of exactly 0 V under standard state conditions, used as the universal reference for measuring electrode potential

standard electrode potential (E_X°)

electrode potential measured under standard conditions (1 bar or 1 atm for gases; 1 *M* for solutes) usually at 298.15 K

Potential, Free Energy, and Equilibrium

By the end of this section, you will be able to:

- Explain the relations between potential, free energy change, and equilibrium constants
- Perform calculations involving the relations between cell potentials, free energy changes, and equilibrium
- Use the Nernst equation to determine cell potentials under nonstandard conditions

So far in this chapter, the relationship between the cell potential and reaction *spontaneity* has been described, suggesting a link to the free energy change for the reaction (see chapter on thermodynamics). The interpretation of potentials as measures of oxidant *strength* was presented, bringing to mind similar measures of acid-base strength as reflected in equilibrium constants (see the chapter on acid-base equilibria). This section provides a summary of the relationships between potential and the related thermodynamic properties ΔG and K .

E° and ΔG°

The standard free energy change of a process, ΔG° , was defined in a previous chapter as the maximum work that could be performed by a system, $w_{\max}$. In the case of a redox reaction taking place within a galvanic cell under standard state conditions, essentially all the work is associated with transferring the electrons from reductant-to-oxidant, w_{elec} :

Equation:

$$\Delta G^\circ = w_{\max} = w_{\text{elec}}$$

The work associated with transferring electrons is determined by the total amount of charge (coulombs) transferred and the cell potential:

Equation:

$$\begin{aligned}\Delta G^\circ = w_{\text{elec}} &= -nFE_{\text{cell}}^\circ \\ \Delta G^\circ &= -nFE_{\text{cell}}^\circ\end{aligned}$$

where n is the number of moles of electrons transferred, F is **Faraday's constant**, and E_{cell}° is the standard cell potential. The relation between free energy change and standard cell potential confirms the sign conventions and spontaneity criteria previously discussed for both of these properties: spontaneous redox reactions exhibit positive potentials and negative free energy changes.

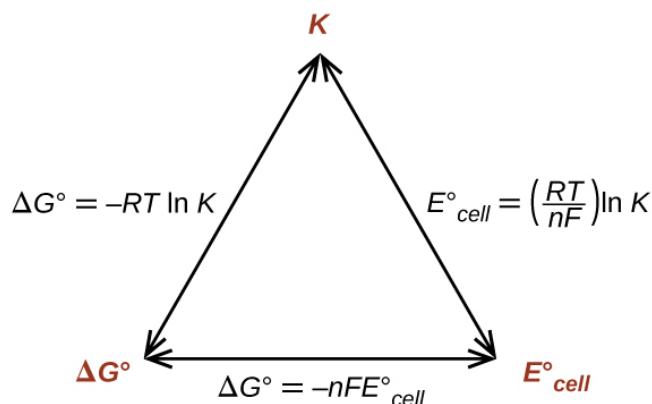
E° and K

Combining a previously derived relation between ΔG° and K (see the chapter on thermodynamics) and the equation above relating ΔG° and E_{cell}° yields the following:

Equation:

$$\begin{aligned}\Delta G^\circ &= -RT \ln K = -nFE_{\text{cell}}^\circ \\ E_{\text{cell}}^\circ &= \left(\frac{RT}{nF}\right) \ln K\end{aligned}$$

This equation indicates redox reactions with large (positive) standard cell potentials will proceed far towards completion, reaching equilibrium when the majority of reactant has been converted to product. A summary of the relations between E° , ΔG° and K is depicted in [\[link\]](#), and a table correlating reaction spontaneity to values of these properties is provided in [\[link\]](#).



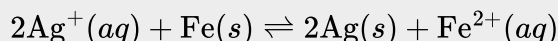
Graphic depicting the relation between three important thermodynamic properties.

K	ΔG°	E°_{cell}	
> 1	< 0	> 0	Reaction is spontaneous under standard conditions Products more abundant at equilibrium
< 1	> 0	< 0	Reaction is non-spontaneous under standard conditions Reactants more abundant at equilibrium
$= 1$	$= 0$	$= 0$	Reaction is at equilibrium under standard conditions Reactants and products equally abundant

Example:
Equilibrium Constants, Standard Cell Potentials, and Standard Free Energy Changes

Use data from [Appendix L](#) to calculate the standard cell potential, standard free energy change, and equilibrium constant for the following reaction at 25 °C. Comment on the spontaneity of the forward reaction and the composition of an equilibrium mixture of reactants and products.

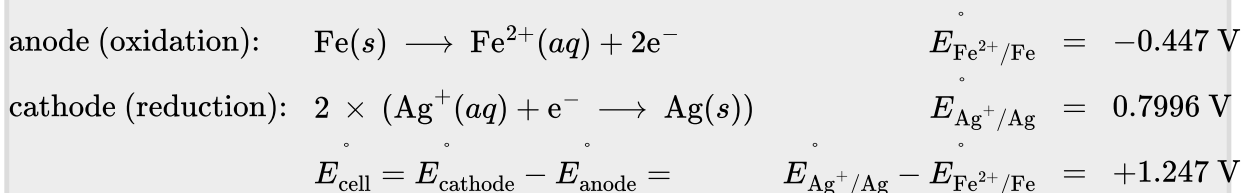
Equation:



Solution

The reaction involves an oxidation-reduction reaction, so the standard cell potential can be calculated using the data in [Appendix L](#).

Equation:



With $n = 2$, the equilibrium constant is then

Equation:

$$\begin{aligned} E_{\text{cell}}^\circ &= \frac{0.0592 \text{ V}}{n} \log K \\ K &= 10^{n \times E_{\text{cell}}^\circ / 0.0592 \text{ V}} \\ K &= 10^{2 \times 1.247 \text{ V} / 0.0592 \text{ V}} \\ K &= 10^{42.128} \\ K &= 1.3 \times 10^{42} \end{aligned}$$

The standard free energy is then

Equation:

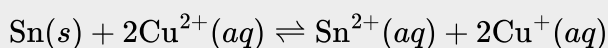
$$\begin{aligned} \Delta G^\circ &= -nFE_{\text{cell}}^\circ \\ \Delta G^\circ &= -2 \times 96,485 \frac{\text{C}}{\text{mol}} \times 1.247 \frac{\text{J}}{\text{C}} = -240.6 \frac{\text{kJ}}{\text{mol}} \end{aligned}$$

The reaction is spontaneous, as indicated by a negative free energy change and a positive cell potential. The K value is very large, indicating the reaction proceeds to near completion to yield an equilibrium mixture containing mostly products.

Check Your Learning

What is the standard free energy change and the equilibrium constant for the following reaction at room temperature? Is the reaction spontaneous?

Equation:



Note:

Answer:

Spontaneous; $n = 2$; $E_{\text{cell}}^{\circ} = +0.291 \text{ V}$; $\Delta G^{\circ} = -56.2 \frac{\text{kJ}}{\text{mol}}$; $K = 6.8 \times 10^9$.

Potentials at Nonstandard Conditions: The Nernst Equation

Most of the redox processes that interest science and society do not occur under standard state conditions, and so the potentials of these systems under nonstandard conditions are a property worthy of attention. Having established the relationship between potential and free energy change in this section, the previously discussed relation between free energy change and reaction mixture composition can be used for this purpose.

Equation:

$$\Delta G = \Delta G^{\circ} + RT \ln Q$$

Notice the reaction quotient, Q , appears in this equation, making the free energy change dependent upon the composition of the reaction mixture. Substituting the equation relating free energy change to cell potential yields the **Nernst equation**:

Equation:

$$-nFE_{\text{cell}} = -nFE_{\text{cell}}^{\circ} + RT \ln Q$$

Equation:

$$E_{\text{cell}} = E_{\text{cell}}^{\circ} - \frac{RT}{nF} \ln Q$$

This equation describes how the potential of a redox system (such as a galvanic cell) varies from its standard state value, specifically, showing it to be a function of the number of electrons transferred, n , the temperature, T , and the reaction mixture composition as reflected in Q . A convenient form of the Nernst equation for most work is one in which values for the fundamental constants (R and F) and a factor converting from natural to base-10 logarithms have been included:

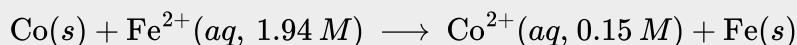
Equation:

$$E_{\text{cell}} = E_{\text{cell}}^{\circ} - \frac{0.0592\text{V}}{n} \ln Q$$

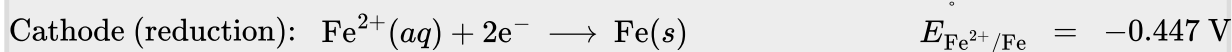
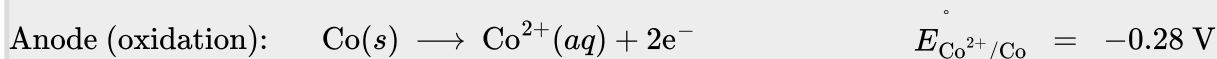
Example:

Predicting Redox Spontaneity Under Nonstandard Conditions

Use the Nernst equation to predict the spontaneity of the redox reaction shown below.

Equation:**Solution**

Collecting information from [Appendix L](#) and the problem,

Equation:

$$E_{\text{cell}}^{\circ} = E_{\text{cathode}}^{\circ} - E_{\text{anode}}^{\circ} = -0.447 \text{ V} - (-0.28 \text{ V}) = -0.17 \text{ V}$$

Notice the negative value of the standard cell potential indicates the process is not spontaneous under standard conditions. Substitution of the Nernst equation terms for the nonstandard conditions yields:

Equation:

$$Q = \frac{[\text{Co}^{2+}]}{[\text{Fe}^{2+}]} = \frac{0.15 M}{1.94 M} = 0.077$$

$$E_{\text{cell}} = E_{\text{cell}}^{\circ} - \frac{0.0592 \text{ V}}{n} \log Q$$

$$E_{\text{cell}} = -0.17 \text{ V} - \frac{0.0592 \text{ V}}{2} \log 0.077$$

$$E_{\text{cell}} = -0.17 \text{ V} + 0.033 \text{ V} = -0.14 \text{ V}$$

The cell potential remains negative (slightly) under the specified conditions, and so the reaction remains nonspontaneous.

Check Your Learning

For the cell schematic below, identify values for n and Q , and calculate the cell potential, E_{cell} .

Equation:**Note:****Answer:**

$n = 6$; $Q = 1440$; $E_{\text{cell}} = +1.97 \text{ V}$, spontaneous.

A **concentration cell** is constructed by connecting two nearly identical half-cells, each based on the same half-reaction and using the same electrode, varying only in the concentration of one redox species. The potential of a concentration cell, therefore, is determined only by the difference in

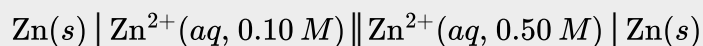
concentration of the chosen redox species. The example problem below illustrates the use of the Nernst equation in calculations involving concentration cells.

Example:

Concentration Cells

What is the cell potential of the concentration cell described by

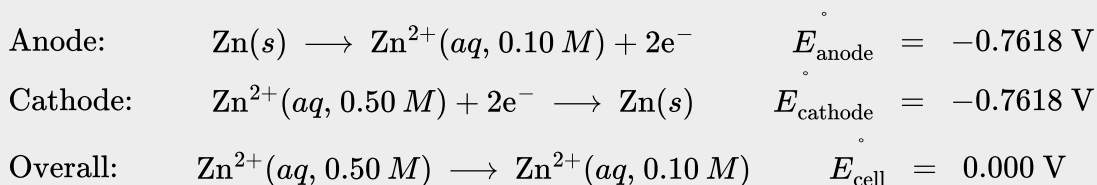
Equation:



Solution

From the information given:

Equation:



Substituting into the Nernst equation,

Equation:

$$E_{\text{cell}} = 0.000\text{ V} - \frac{0.0592\text{ V}}{2} \log \frac{0.10}{0.50} = +0.021\text{ V}$$

The positive value for cell potential indicates the overall cell reaction (see above) is spontaneous. This spontaneous reaction is one in which the zinc ion concentration in the cathode falls (it is reduced to elemental zinc) while that in the anode rises (it is produced by oxidation of the zinc anode). A greater driving force for zinc reduction is present in the cathode, where the zinc(II) ion concentration is greater ($E_{\text{cathode}} > E_{\text{anode}}$).

Check Your Learning

The concentration cell above was allowed to operate until the cell reaction reached equilibrium. What are the cell potential and the concentrations of zinc(II) in each half-cell for the cell now?

Note:

Answer:

$$E_{\text{cell}} = 0.000\text{ V}; [\text{Zn}^{2+}]_{\text{cathode}} = [\text{Zn}^{2+}]_{\text{anode}} = 0.30\text{ M}$$

Key Concepts and Summary

Potential is a thermodynamic quantity reflecting the intrinsic driving force of a redox process, and it is directly related to the free energy change and equilibrium constant for the process. For redox processes taking place in electrochemical cells, the maximum (electrical) work done by the system is easily computed from the cell potential and the reaction stoichiometry and is equal to the free energy change for the process. The equilibrium constant for a redox reaction is logarithmically related to the reaction's cell potential, with larger (more positive) potentials indicating reactions with greater driving force that equilibrate when the reaction has proceeded far towards completion (large value of K). Finally, the potential of a redox process varies with the composition of the reaction mixture, being related to the reactions standard potential and the value of its reaction quotient, Q , as described by the Nernst equation.

Key Equations

- $E_{\text{cell}}^{\circ} = \frac{RT}{nF} \ln K$
- $E_{\text{cell}}^{\circ} = \frac{0.0257 \text{ V}}{n} \ln K = \frac{0.0592 \text{ V}}{n} \log K$ (at 298.15 K)
- $E_{\text{cell}} = E_{\text{cell}}^{\circ} - \frac{RT}{nF} \ln Q$ (Nernst equation)
- $E_{\text{cell}} - \frac{0.0592 \text{ V}}{n} \log Q$ (at 298.15 K)
- $\Delta G = -nFE_{\text{cell}}^{\circ}$
- $\Delta G^{\circ} = -nFE_{\text{cell}}^{\circ}$
- $w_{\text{ele}} = w_{\text{max}} = -nFE_{\text{cell}}^{\circ}$

Chemistry End of Chapter Exercises

Exercise:

Problem:

For each pair of standard cell potential and electron stoichiometry values below, calculate a corresponding standard free energy change (kJ).

(a) 0.000 V, $n = 2$

(b) +0.434 V, $n = 2$

(c) -2.439 V, $n = 1$

Solution:

(a) 0 kJ/mol; (b) -83.7 kJ/mol; (c) +235.3 kJ/mol

Exercise:

Problem:

For each pair of standard free energy change and electron stoichiometry values below, calculate a corresponding standard cell potential.

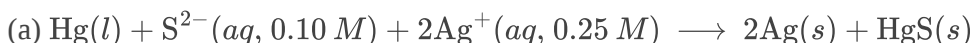
(a) 12 kJ/mol, $n = 3$

(b) -45 kJ/mol , $n = 1$

Exercise:

Problem:

Determine the standard cell potential and the cell potential under the stated conditions for the electrochemical reactions described here. State whether each is spontaneous or nonspontaneous under each set of conditions at 298.15 K.



(b) The cell made from an anode half-cell consisting of an aluminum electrode in 0.015 *M* aluminum nitrate solution and a cathode half-cell consisting of a nickel electrode in 0.25 *M* nickel(II) nitrate solution.

(c) The cell made of a half-cell in which 1.0 *M* aqueous bromide is oxidized to 0.11 *M* bromine ion and a half-cell in which aluminum ion at 0.023 *M* is reduced to aluminum metal.

Solution:

(a) standard cell potential: 1.50 V, spontaneous; cell potential under stated conditions: 1.43 V, spontaneous; (b) standard cell potential: 1.405 V, spontaneous; cell potential under stated conditions: 1.423 V, spontaneous; (c) standard cell potential: -2.749 V , nonspontaneous; cell potential under stated conditions: -2.757 V , nonspontaneous

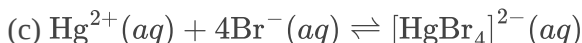
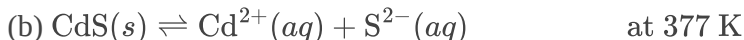
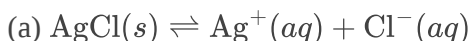
Exercise:

Problem: Determine ΔG and ΔG° for each of the reactions in the previous problem.

Exercise:

Problem:

Use the data in [Appendix L](#) to calculate equilibrium constants for the following reactions. Assume 298.15 K if no temperature is given.



Solution:

(a) 1.7×10^{-10} ; (b) 2.6×10^{-21} ; (c) 4.693×10^{21} ; (d) 1.0×10^{-14}

Glossary

concentration cell

galvanic cell comprising half-cells of identical composition but for the concentration of one redox reactant or product

Faraday's constant (F)

charge on 1 mol of electrons; $F = 96,485 \text{ C/mol e}^-$

Nernst equation

relating the potential of a redox system to its composition

Batteries and Fuel Cells

By the end of this section, you will be able to:

- Describe the electrochemistry associated with several common batteries
- Distinguish the operation of a fuel cell from that of a battery

There are many technological products associated with the past two centuries of electrochemistry research, none more immediately obvious than the battery. A **battery** is a galvanic cell that has been specially designed and constructed in a way that best suits its intended use as a source of electrical power for specific applications. Among the first successful batteries was the *Daniell cell*, which relied on the spontaneous oxidation of zinc by copper(II) ions ([link](#)):

Equation:

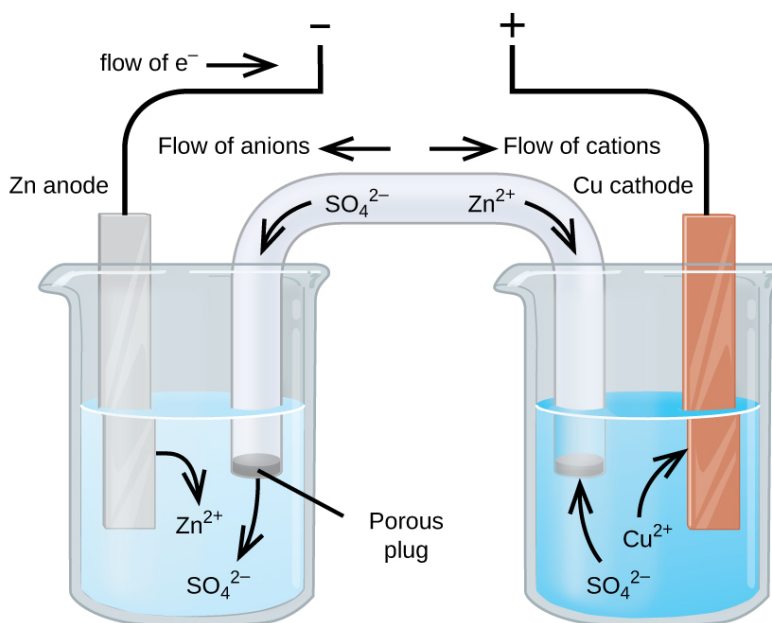
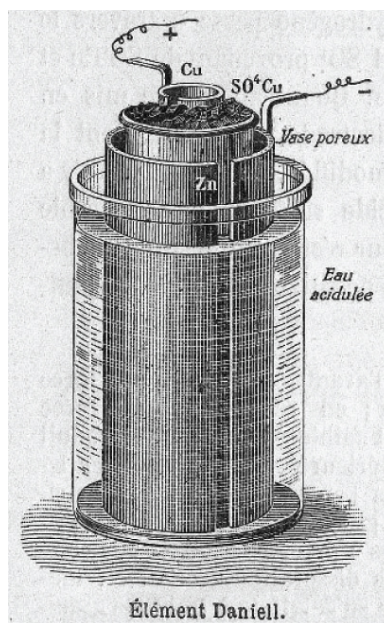
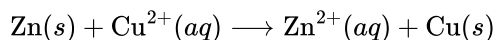


Illustration of a Daniell cell taken from a 1904 journal publication (left) along with a simplified illustration depicting the electrochemistry of the cell (right). The 1904 design used a porous clay pot to both contain one of the half-cell's content and to serve as a salt bridge to the other half-cell.

Modern batteries exist in a multitude of forms to accommodate various applications, from tiny button batteries that provide the modest power needs of a wristwatch to the very large batteries used to supply backup energy to municipal power grids. Some batteries are designed for single-use applications and cannot be recharged (**primary cells**), while others are based on conveniently reversible cell reactions that allow recharging by an external power source (**secondary cells**). This section will provide a summary of the basic electrochemical aspects of several batteries familiar to most consumers, and will introduce a related electrochemical device called a *fuel cell* that can offer improved performance in certain applications.

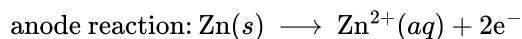
Note:

Visit this [site](#) to learn more about batteries.

Single-Use Batteries

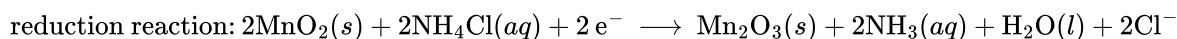
A common primary battery is the **dry cell**, which uses a zinc can as both container and anode (“−” terminal) and a graphite rod as the cathode (“+” terminal). The Zn can is filled with an electrolyte paste containing manganese(IV) oxide, zinc(II) chloride, ammonium chloride, and water. A graphite rod is immersed in the electrolyte paste to complete the cell. The spontaneous cell reaction involves the oxidation of zinc:

Equation:



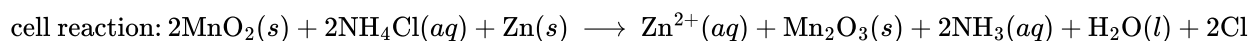
and the reduction of manganese(IV)

Equation:

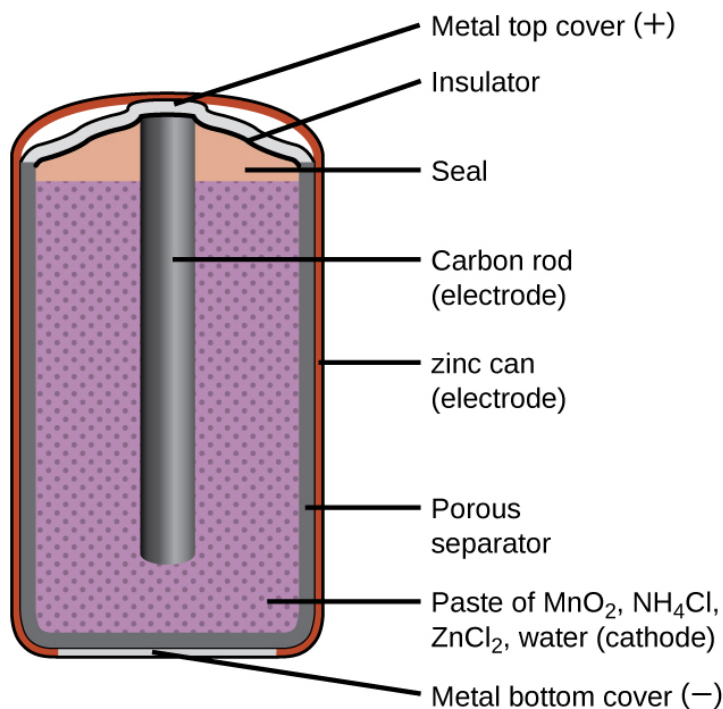


which together yield the cell reaction:

Equation:



The voltage (*cell potential*) of a dry cell is approximately 1.5 V. Dry cells are available in various sizes (e.g., D, C, AA, AAA). All sizes of dry cells comprise the same components, and so they exhibit the same voltage, but larger cells contain greater amounts of the redox reactants and therefore are capable of transferring correspondingly greater amounts of charge. Like other galvanic cells, dry cells may be connected in series to yield batteries with greater voltage outputs, if needed.

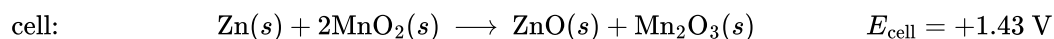
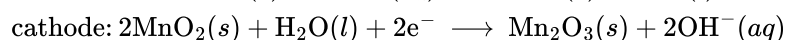
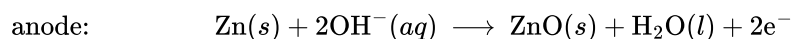


A schematic diagram shows a typical dry cell.

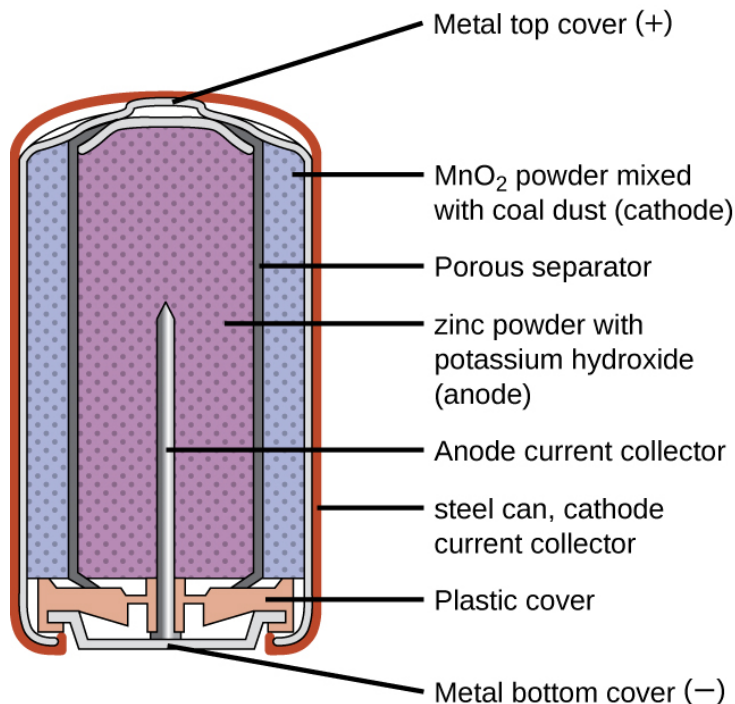
Note:

Visit this [site](#) to learn more about zinc-carbon batteries.

Alkaline batteries ([link](#)) were developed in the 1950s to improve on the performance of the dry cell, and they were designed around the same redox couples. As their name suggests, these types of batteries use alkaline electrolytes, often potassium hydroxide. The reactions are

Equation:

An alkaline battery can deliver about three to five times the energy of a zinc-carbon dry cell of similar size. Alkaline batteries are prone to leaking potassium hydroxide, so they should be removed from devices for long-term storage. While some alkaline batteries are rechargeable, most are not. Attempts to recharge an alkaline battery that is not rechargeable often leads to rupture of the battery and leakage of the potassium hydroxide electrolyte.



Alkaline batteries were designed as improved replacements for zinc-carbon (dry cell) batteries.

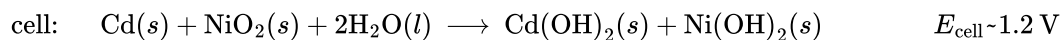
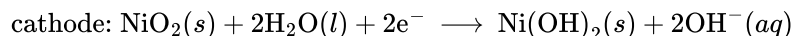
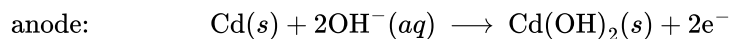
Note:

Visit this [site](#) to learn more about alkaline batteries.

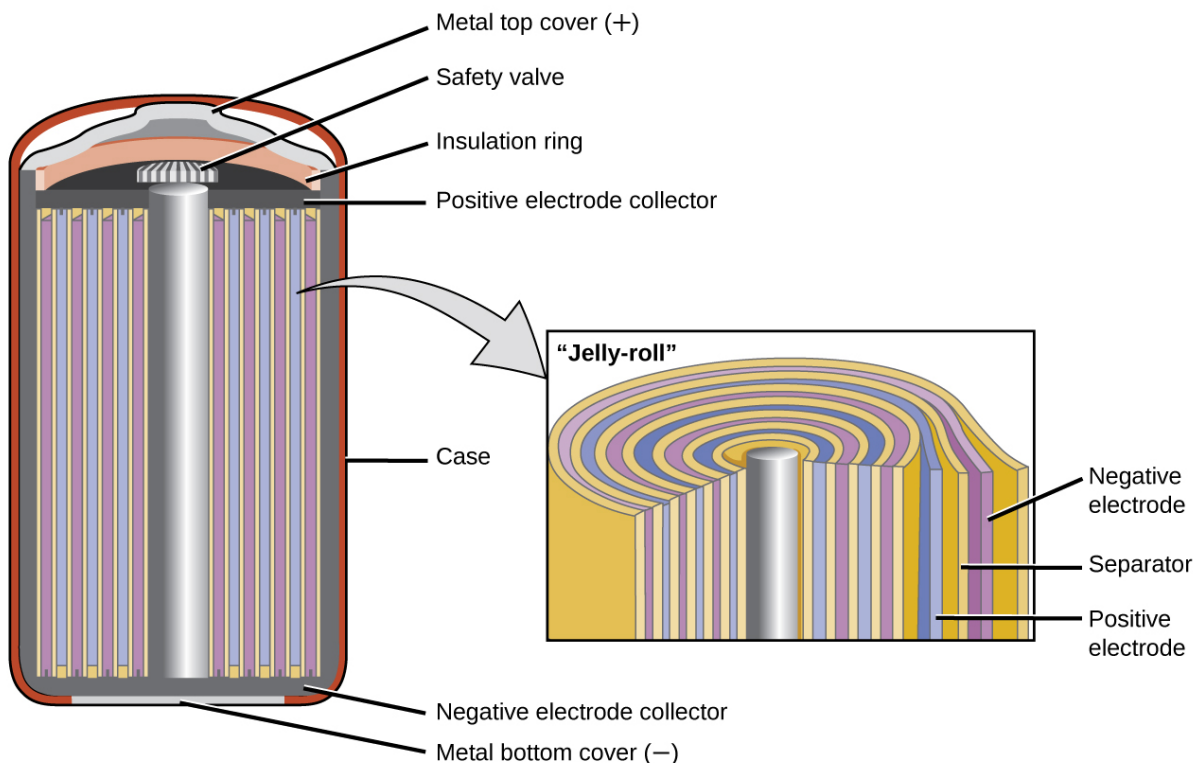
Rechargeable (Secondary) Batteries

Nickel-cadmium, or NiCd, batteries ([\[link\]](#)) consist of a nickel-plated cathode, cadmium-plated anode, and a potassium hydroxide electrolyte. The positive and negative plates, which are prevented from shorting by the separator, are rolled together and put into the case. This is a “jelly-roll” design and allows the NiCd cell to deliver much more current than a similar-sized alkaline battery. The reactions are

Equation:



When properly treated, a NiCd battery can be recharged about 1000 times. Cadmium is a toxic heavy metal so NiCd batteries should never be ruptured or incinerated, and they should be disposed of in accordance with relevant toxic waste guidelines.



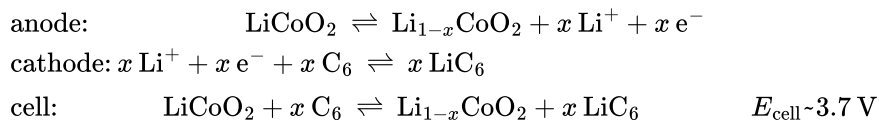
NiCd batteries use a “jelly-roll” design that significantly increases the amount of current the battery can deliver as compared to a similar-sized alkaline battery.

Note:

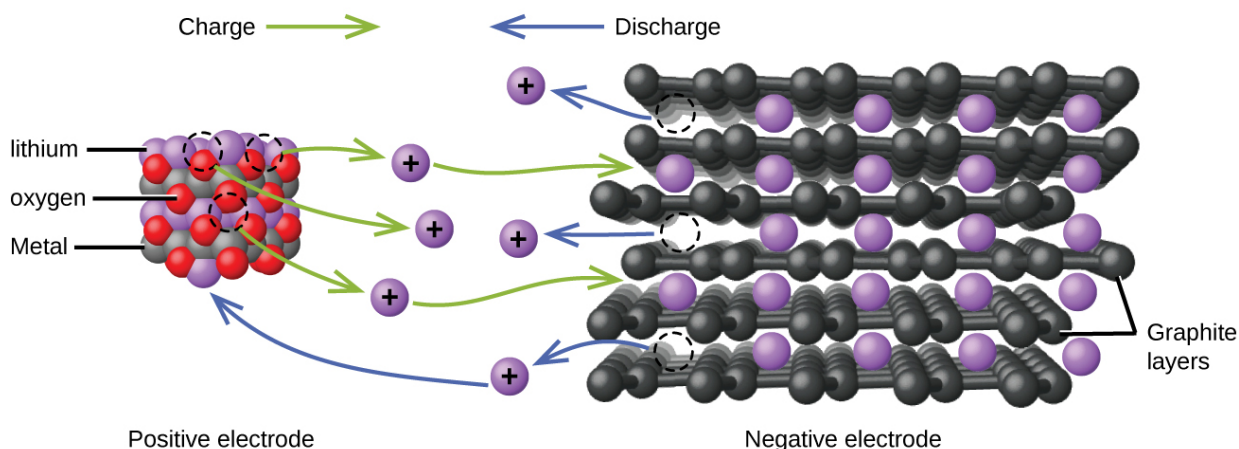
Visit this [site](#) for more information about nickel cadmium rechargeable batteries.

Lithium ion batteries ([link](#)) are among the most popular rechargeable batteries and are used in many portable electronic devices. The reactions are

Equation:



The variable stoichiometry of the cell reaction leads to variation in cell voltages, but for typical conditions, x is usually no more than 0.5 and the cell voltage is approximately 3.7 V. Lithium batteries are popular because they can provide a large amount current, are lighter than comparable batteries of other types, produce a nearly constant voltage as they discharge, and only slowly lose their charge when stored.



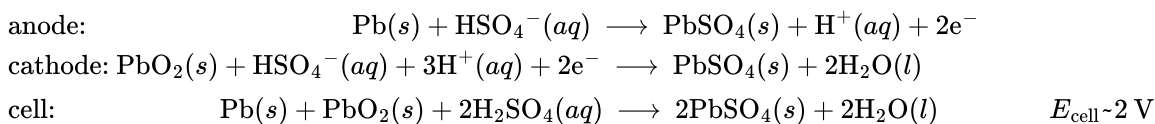
In a lithium ion battery, charge flows as the lithium ions are transferred between the anode and cathode.

Note:

Visit this [site](#) for more information about lithium ion batteries.

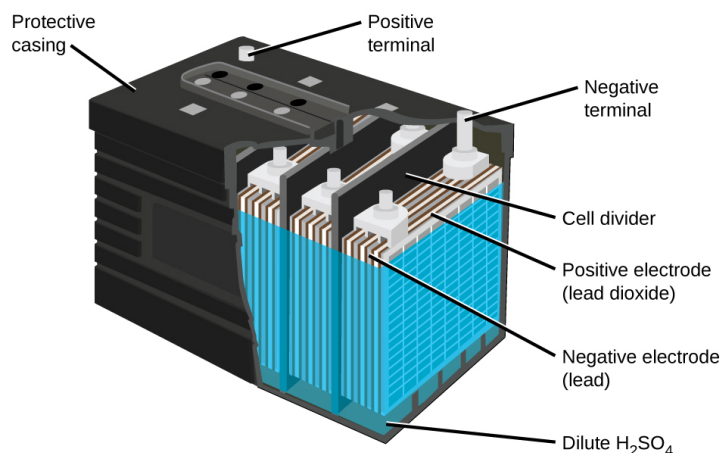
The **lead acid battery** ([link](#)) is the type of secondary battery commonly used in automobiles. It is inexpensive and capable of producing the high current required by automobile starter motors. The reactions for a lead acid battery are

Equation:



Each cell produces 2 V, so six cells are connected in series to produce a 12-V car battery. Lead acid batteries are heavy and contain a caustic liquid electrolyte, $\text{H}_2\text{SO}_4(aq)$, but are often still the battery of choice because of their

high current density. Since these batteries contain a significant amount of lead, they must always be disposed of properly.



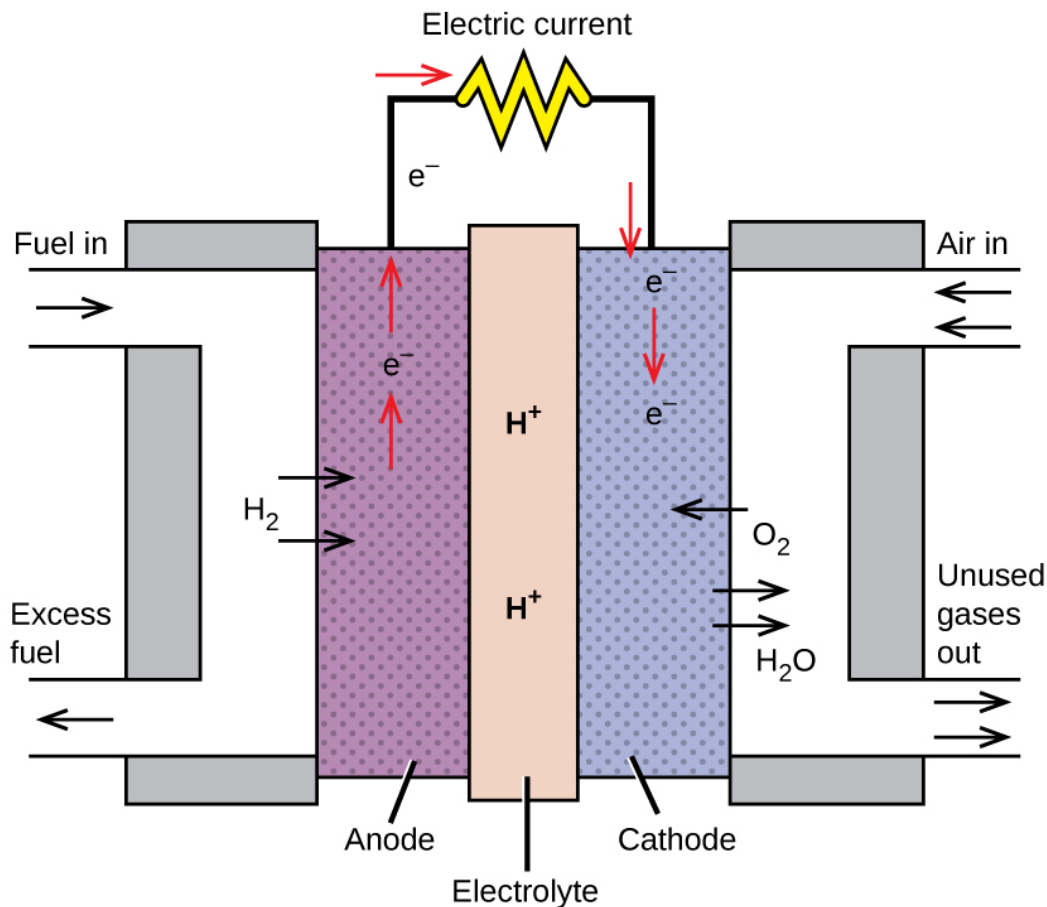
The lead acid battery in your automobile consists of six cells connected in series to give 12 V.

Note:

Visit this [site](#) for more information about lead acid batteries.

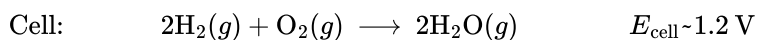
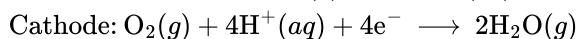
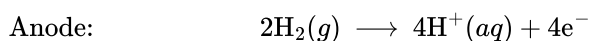
Fuel Cells

A **fuel cell** is a galvanic cell that uses traditional combustible fuels, most often hydrogen or methane, that are continuously fed into the cell along with an oxidant. (An alternative, but not very popular, name for a fuel cell is a *flow battery*.) Within the cell, fuel and oxidant undergo the same redox chemistry as when they are combusted, but via a catalyzed electrochemical that is significantly more efficient. For example, a typical hydrogen fuel cell uses graphite electrodes embedded with platinum-based catalysts to accelerate the two half-cell reactions:



In this hydrogen fuel cell, oxygen from the air reacts with hydrogen, producing water and electricity.

Equation:



These types of fuel cells generally produce voltages of approximately 1.2 V. Compared to an internal combustion engine, the energy efficiency of a fuel cell using the same redox reaction is typically more than double (~20%–25% for an engine versus ~50%–75% for a fuel cell). Hydrogen fuel cells are commonly used on extended space missions, and prototypes for personal vehicles have been developed, though the technology remains relatively immature.

Note:

Check out this [link](#) to learn more about fuel cells.

Key Concepts and Summary

Galvanic cells designed specifically to function as electrical power supplies are called batteries. A variety of both single-use batteries (primary cells) and rechargeable batteries (secondary cells) are commercially available to serve a variety of applications, with important specifications including voltage, size, and lifetime. Fuel cells, sometimes called flow batteries, are devices that harness the energy of spontaneous redox reactions normally associated with combustion processes. Like batteries, fuel cells enable the reaction's electron transfer via an external circuit, but they require continuous input of the redox reactants (fuel and oxidant) from an external reservoir. Fuel cells are typically much more efficient in converting the energy released by the reaction to useful work in comparison to internal combustion engines.

Chemistry End of Chapter Exercises

Exercise:

Problem:

Consider a battery made from one half-cell that consists of a copper electrode in 1 M CuSO₄ solution and another half-cell that consists of a lead electrode in 1 M Pb(NO₃)₂ solution.

- What is the standard cell potential for the battery?
- What are the reactions at the anode, cathode, and the overall reaction?
- Most devices designed to use dry-cell batteries can operate between 1.0 and 1.5 V. Could this cell be used to make a battery that could replace a dry-cell battery? Why or why not.
- Suppose sulfuric acid is added to the half-cell with the lead electrode and some PbSO₄(s) forms. Would the cell potential increase, decrease, or remain the same?

Exercise:

Problem: Consider a battery with the overall reaction: $\text{Cu}(s) + 2\text{Ag}^+(aq) \longrightarrow 2\text{Ag}(s) + \text{Cu}^{2+}(aq)$.

- What is the reaction at the anode and cathode?
- A battery is "dead" when its cell potential is zero. What is the value of Q when this battery is dead?
- If a particular dead battery was found to have $[\text{Cu}^{2+}] = 0.11 \text{ M}$, what was the concentration of silver ion?

Solution:

$$\begin{array}{ll} \text{(a) anode: } \text{Cu}(s) \longrightarrow \text{Cu}^{2+}(aq) + 2e^- & E_{\text{anode}} = 0.34 \text{ V} \\ \text{cathode: } 2 \times (\text{Ag}^+(aq) + e^- \longrightarrow \text{Ag}(s)) & E_{\text{cathode}} = 0.7996 \text{ V} \end{array} \quad ; \text{ (b) } 3.5 \times 10^{15}; \text{ (c) } 5.6 \times 10^{-9} \text{ M}$$

Exercise:

Problem: Why do batteries go dead, but fuel cells do not?

Solution:

Batteries are self-contained and have a limited supply of reagents to expend before going dead. Alternatively, battery reaction byproducts accumulate and interfere with the reaction. Because a fuel cell is constantly resupplied with reactants and products are expelled, it can continue to function as long as reagents are supplied.

Exercise:

Problem:

Use the Nernst equation to explain the drop in voltage observed for some batteries as they discharge.

Exercise:**Problem:**

Using the information thus far in this chapter, explain why battery-powered electronics perform poorly in low temperatures.

Solution:

E_{cell} , as described in the Nernst equation, has a term that is directly proportional to temperature. At low temperatures, this term is decreased, resulting in a lower cell voltage provided by the battery to the device—the same effect as a battery running dead.

Glossary

alkaline battery

primary battery similar to a *dry cell* that uses an alkaline (often potassium hydroxide) electrolyte; designed to be an improved replacement for the dry cell, but with more energy storage and less electrolyte leakage than typical dry cell

battery

single or series of galvanic cells designed for use as a source of electrical power

dry cell

primary battery, also called a zinc-carbon battery, based on the spontaneous oxidation of zinc by manganese(IV)

fuel cell

devices similar to galvanic cells that require a continuous feed of redox reactants; also called a *flow battery*

lead acid battery

rechargeable battery commonly used in automobiles; it typically comprises six galvanic cells based on Pb half-reactions in acidic solution

lithium ion battery

widely used rechargeable battery commonly used in portable electronic devices, based on lithium ion transfer between the anode and cathode

nickel-cadmium battery

rechargeable battery based on Ni/Cd half-cells with applications similar to those of lithium ion batteries

primary cell

nonrechargeable battery, suitable for single use only

secondary cell

battery designed to allow recharging

Corrosion

By the end of this section, you will be able to:

- Define corrosion
- List some of the methods used to prevent or slow corrosion

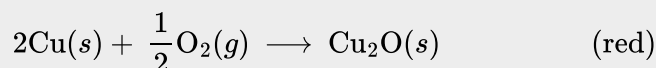
Corrosion is usually defined as the degradation of metals by a naturally occurring electrochemical process. The formation of rust on iron, tarnish on silver, and the blue-green patina that develops on copper are all examples of corrosion. The total cost of corrosion remediation in the United States is significant, with estimates in excess of half a trillion dollars a year.

Note:

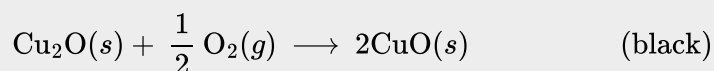
Statue of Liberty: Changing Colors

The Statue of Liberty is a landmark every American recognizes. The Statue of Liberty is easily identified by its height, stance, and unique blue-green color ([link](#)). When this statue was first delivered from France, its appearance was not green. It was brown, the color of its copper “skin.” So how did the Statue of Liberty change colors? The change in appearance was a direct result of corrosion. The copper that is the primary component of the statue slowly underwent oxidation from the air. The oxidation-reduction reactions of copper metal in the environment occur in several steps. Copper metal is oxidized to copper(I) oxide (Cu_2O), which is red, and then to copper(II) oxide, which is black

Equation:



Equation:



Coal, which was often high in sulfur, was burned extensively in the early part of the last century. As a result, atmospheric sulfur trioxide, carbon dioxide, and water all reacted with the CuO

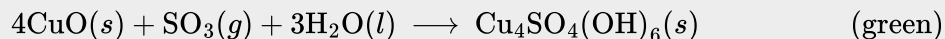
Equation:



Equation:



Equation:



These three compounds are responsible for the characteristic blue-green patina seen on the Statue of Liberty (and other outdoor copper structures). Fortunately, formation of patina creates a protective layer on the copper surface, preventing further corrosion of the underlying copper. The formation of the protective layer is called *passivation*, a phenomenon discussed further in another chapter of this text.



(a)

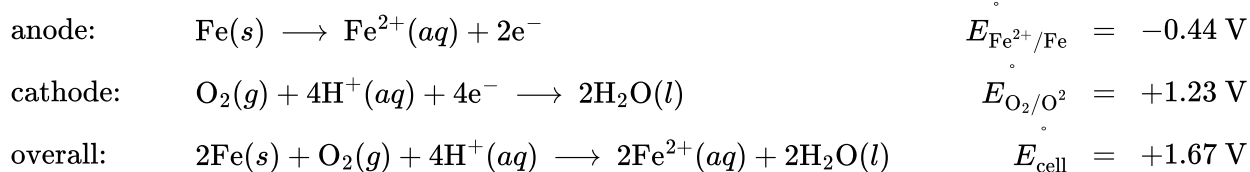


(b)

(a) The Statue of Liberty is covered with a copper skin, and was originally brown, as shown in this painting. (b) Exposure to the elements has resulted in the formation of the blue-green patina seen today.

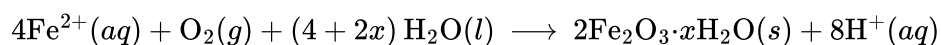
Perhaps the most familiar example of corrosion is the formation of rust on iron. Iron will rust when it is exposed to oxygen and water. Rust formation involves the creation of a galvanic cell at an iron surface, as illustrated in [\[link\]](#). The relevant redox reactions are described by the following equations:

Equation:

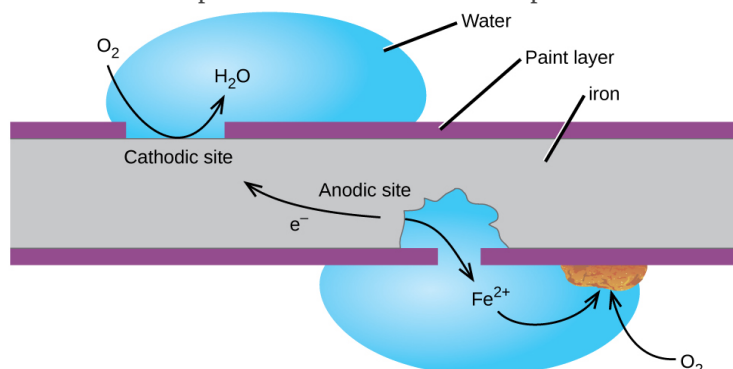


Further reaction of the iron(II) product in humid air results in the production of an iron(III) oxide hydrate known as rust:

Equation:



The stoichiometry of the hydrate varies, as indicated by the use of x in the compound formula. Unlike the patina on copper, the formation of rust does not create a protective layer and so corrosion of the iron continues as the rust flakes off and exposes fresh iron to the atmosphere.



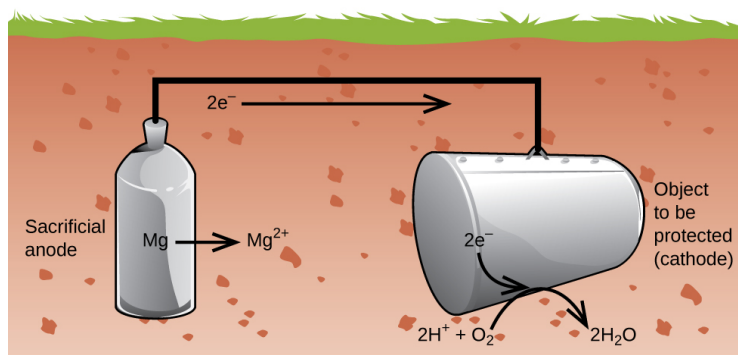
Corrosion can occur when a painted iron or steel surface is exposed to the environment by a scratch through the paint. A galvanic cell results that may be approximated by the simplified cell schematic $Fe(s) | Fe^{2+}(aq) || O_2(aq), H_2O(l) | Fe(s)$.

One way to keep iron from corroding is to keep it painted. The layer of paint prevents the water and oxygen necessary for rust formation from coming into contact with the iron. As long as the paint remains intact, the iron is protected from corrosion.

Other strategies include alloying the iron with other metals. For example, stainless steel is an alloy of iron containing a small amount of chromium. The chromium tends to collect near the surface, where it corrodes and forms a passivating oxide layer that protects the iron.

Iron and other metals may also be protected from corrosion by **galvanization**, a process in which the metal to be protected is coated with a layer of a more readily oxidized metal, usually zinc. When the zinc layer is intact, it prevents air from contacting the underlying iron and thus prevents corrosion. If the zinc layer is breached by either corrosion or mechanical abrasion, the iron may still be protected from corrosion by a *cathodic protection* process, which is described in the next paragraph.

Another important way to protect metal is to make it the cathode in a galvanic cell. This is **cathodic protection** and can be used for metals other than just iron. For example, the rusting of underground iron storage tanks and pipes can be prevented or greatly reduced by connecting them to a more active metal such as zinc or magnesium ([link](#)). This is also used to protect the metal parts in water heaters. The more active metals (lower reduction potential) are called **sacrificial anodes** because as they get used up as they corrode (oxidize) at the anode. The metal being protected serves as the cathode for the reduction of oxygen in air, and so it simply serves to conduct (not react with) the electrons being transferred. When the anodes are properly monitored and periodically replaced, the useful lifetime of the iron storage tank can be greatly extended.



Cathodic protection is a useful approach to electrochemically preventing corrosion of underground storage tanks.

Key Concepts and Summary

Spontaneous oxidation of metals by natural electrochemical processes is called corrosion, familiar examples including the rusting of iron and the tarnishing of silver. Corrosion process involve the creation of a galvanic cell in which different sites on the metal object function as anode and cathode, with the corrosion taking place at the anodic site. Approaches to preventing corrosion of metals include use of a protective coating of zinc (galvanization) and the use of sacrificial anodes connected to the metal object (cathodic protection).

Chemistry End of Chapter Exercises

Exercise:

Problem: Which member of each pair of metals is more likely to corrode (oxidize)?

- (a) Mg or Ca
- (b) Au or Hg
- (c) Fe or Zn
- (d) Ag or Pt

Exercise:

Problem:

Consider the following metals: Ag, Au, Mg, Ni, and Zn. Which of these metals could be used as a sacrificial anode in the cathodic protection of an underground steel storage tank? Steel is an alloy composed mostly of iron, so use -0.447 V as the standard reduction potential for steel.

Solution:

Mg and Zn

Exercise:**Problem:**

Aluminum ($E_{\text{Al}^{3+}/\text{Al}}^\circ = -2.07 \text{ V}$) is more easily oxidized than iron ($E_{\text{Fe}^{3+}/\text{Fe}}^\circ = -0.477 \text{ V}$), and yet when both are exposed to the environment, untreated aluminum has very good corrosion resistance while the corrosion resistance of untreated iron is poor. What might explain this observation?

Exercise:**Problem:**

If a sample of iron and a sample of zinc come into contact, the zinc corrodes but the iron does not. If a sample of iron comes into contact with a sample of copper, the iron corrodes but the copper does not. Explain this phenomenon.

Solution:

Both examples involve cathodic protection. The (sacrificial) anode is the metal that corrodes (oxidizes or reacts). In the case of iron (-0.447 V) and zinc (-0.7618 V), zinc has a more negative standard reduction potential and so serves as the anode. In the case of iron and copper (0.34 V), iron has the smaller standard reduction potential and so corrodes (serves as the anode).

Exercise:**Problem:**

Suppose you have three different metals, A, B, and C. When metals A and B come into contact, B corrodes and A does not corrode. When metals A and C come into contact, A corrodes and C does not corrode. Based on this information, which metal corrodes and which metal does not corrode when B and C come into contact?

Exercise:

Problem: Why would a sacrificial anode made of lithium metal be a bad choice

Solution:

While the reduction potential of lithium would make it capable of protecting the other metals, this high potential is also indicative of how reactive lithium is; it would have a spontaneous reaction with most substances. This means that the lithium would react quickly with other substances, even those that would not oxidize the metal it is attempting to protect. Reactivity like this means the sacrificial anode would be depleted rapidly and need to be replaced frequently. (Optional additional reason: fire hazard in the presence of water.)

Glossary

cathodic protection

approach to preventing corrosion of a metal object by connecting it to a *sacrificial anode* composed of a more readily oxidized metal

corrosion

degradation of metal via a natural electrochemical process

galvanization

method of protecting iron or similar metals from corrosion by coating with a thin layer of more easily oxidized zinc.

sacrificial anode

electrode constructed from an easily oxidized metal, often magnesium or zinc, used to prevent corrosion of metal objects via cathodic protection

salt bridge

tube filled with inert electrolyte solution

Electrolysis

By the end of this section you will be able to:

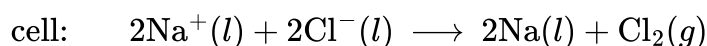
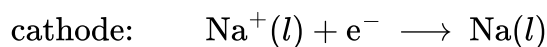
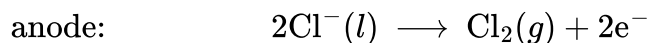
- Describe the process of electrolysis
- Compare the operation of electrolytic cells with that of galvanic cells
- Perform stoichiometric calculations for electrolytic processes

Electrochemical cells in which spontaneous redox reactions take place (*galvanic cells*) have been the topic of discussion so far in this chapter. In these cells, *electrical work is done by a redox system on its surroundings* as electrons produced by the redox reaction are transferred through an external circuit. This final section of the chapter will address an alternative scenario in which *an external circuit does work on a redox system* by imposing a voltage sufficient to drive an otherwise nonspontaneous reaction, a process known as **electrolysis**. A familiar example of electrolysis is recharging a battery, which involves use of an external power source to drive the spontaneous (discharge) cell reaction in the reverse direction, restoring to some extent the composition of the half-cells and the voltage of the battery. Perhaps less familiar is the use of electrolysis in the refinement of metallic ores, the manufacture of commodity chemicals, and the *electroplating* of metallic coatings on various products (e.g., jewelry, utensils, auto parts). To illustrate the essential concepts of electrolysis, a few specific processes will be considered.

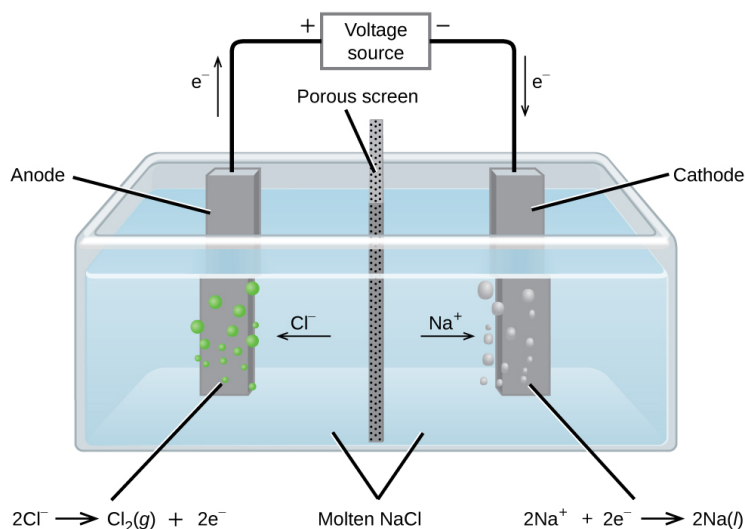
The Electrolysis of Molten Sodium Chloride

Metallic sodium, Na, and chlorine gas, Cl₂, are used in numerous applications, and their industrial production relies on the large-scale electrolysis of molten sodium chloride, NaCl(l). The industrial process typically uses a *Downs cell* similar to the simplified illustration shown in [\[link\]](#). The reactions associated with this process are:

Equation:



The cell potential for the above process is negative, indicating the reaction as written (decomposition of liquid NaCl) is not spontaneous. To force this reaction, *a positive potential of magnitude greater than the negative cell potential* must be applied to the cell.

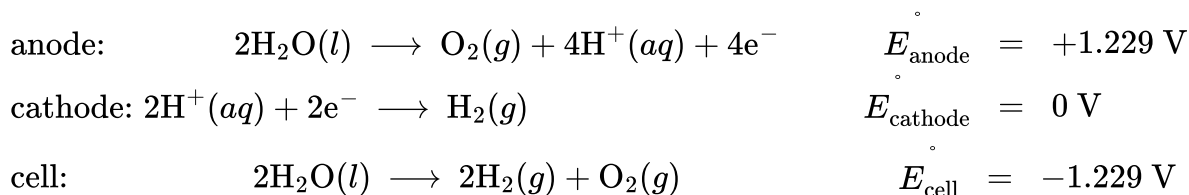


Cells of this sort (a cell for the electrolysis of molten sodium chloride) are used in the *Downs process* for production of sodium and chlorine, and they typically use iron cathodes and carbon anodes.

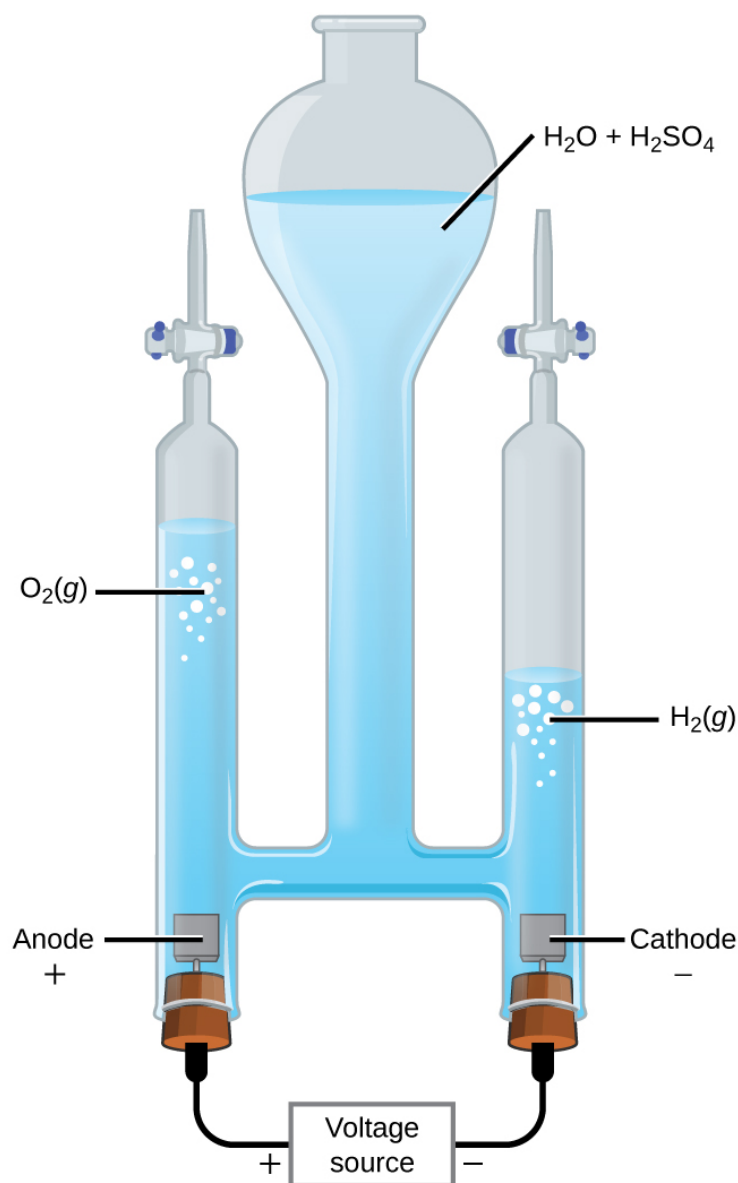
The Electrolysis of Water

Water may be electrolytically decomposed in a cell similar to the one illustrated in [\[link\]](#). To improve electrical conductivity without introducing a different redox species, the hydrogen ion concentration of the water is typically increased by addition of a strong acid. The redox processes associated with this cell are

Equation:



Again, the cell potential as written is negative, indicating a nonspontaneous cell reaction that must be driven by imposing a cell voltage greater than +1.229 V. Keep in mind that *standard* electrode potentials are used to inform thermodynamic predictions here, though the cell is *not* operating under standard state conditions. Therefore, at best, calculated cell potentials should be considered ballpark estimates.

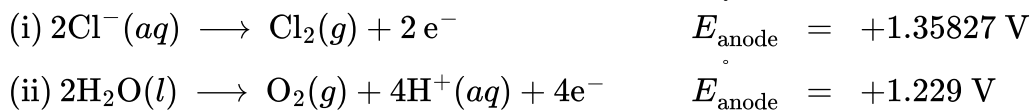


The electrolysis of water produces stoichiometric amounts of oxygen gas at the anode and hydrogen at the anode.

The Electrolysis of Aqueous Sodium Chloride

When aqueous solutions of ionic compounds are electrolyzed, the anode and cathode half-reactions may involve the electrolysis of either water species (H_2O , H^+ , OH^-) or solute species (the cations and anions of the compound). As an example, the electrolysis of aqueous sodium chloride could involve either of these two anode reactions:

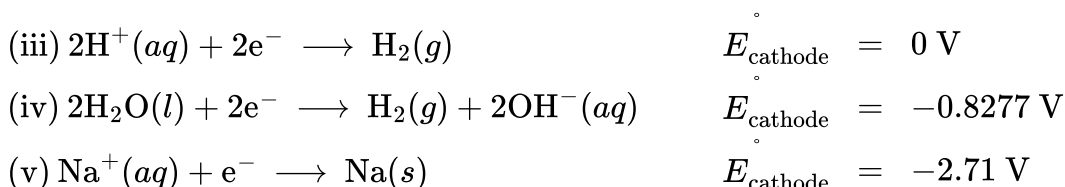
Equation:



The standard electrode (*reduction*) potentials of these two half-reactions indicate water may be *oxidized* at a less negative/more positive potential (−1.229 V) than chloride ion (−1.358 V). Thermodynamics thus predicts that water would be more readily oxidized, though in practice it is observed that both water and chloride ion are oxidized under typical conditions, producing a mixture of oxygen and chlorine gas.

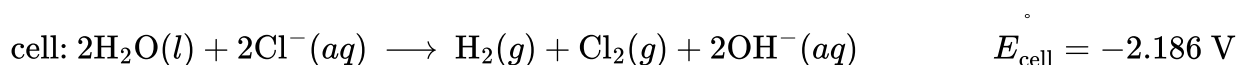
Turning attention to the cathode, the possibilities for reduction are:

Equation:



Comparison of these *standard* half-reaction potentials suggests the reduction of hydrogen ion is thermodynamically favored. However, in a neutral aqueous sodium chloride solution, the concentration of hydrogen ion is far below the standard state value of 1 M (approximately 10^{-7} M), and so the observed cathode reaction is actually reduction of water. The net cell reaction in this case is then

Equation:

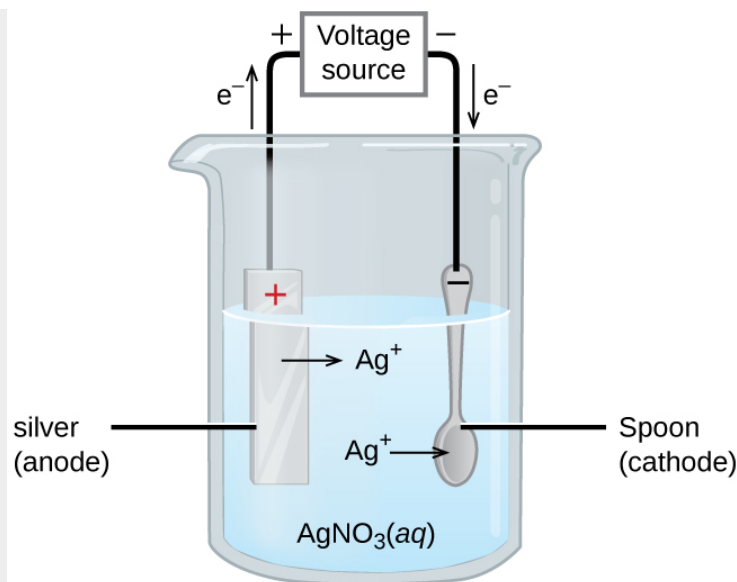


This electrolysis reaction is part of the *chlor-alkali process* used by industry to produce chlorine and sodium hydroxide (lye).

Note:

Electroplating

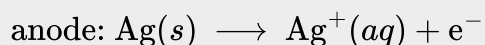
An important use for electrolytic cells is in **electroplating**. Electroplating results in a thin coating of one metal on top of a conducting surface. Reasons for electroplating include making the object more corrosion resistant, strengthening the surface, producing a more attractive finish, or for purifying metal. The metals commonly used in electroplating include cadmium, chromium, copper, gold, nickel, silver, and tin. Common consumer products include silver-plated or gold-plated tableware, chrome-plated automobile parts, and jewelry. The silver plating of eating utensils is used here to illustrate the process. ([link](#)).



This schematic shows an electrolytic cell for silver plating eating utensils.

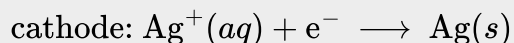
In the figure, the anode consists of a silver electrode, shown on the left. The cathode is located on the right and is the spoon, which is made from inexpensive metal. Both electrodes are immersed in a solution of silver nitrate. Applying a sufficient potential results in the oxidation of the silver anode

Equation:



and reduction of silver ion at the (spoon) cathode:

Equation:



The net result is the transfer of silver metal from the anode to the cathode. Several experimental factors must be carefully controlled to obtain high-quality silver coatings, including the exact composition of the electrolyte solution, the cell voltage applied, and the rate of the electrolysis reaction (electrical current).

Quantitative Aspects of Electrolysis

Electrical current is defined as the rate of flow for any charged species. Most relevant to this discussion is the flow of electrons. Current is measured in a composite unit called an ampere, defined as one coulomb per second ($A = 1 \text{ C/s}$). The charge transferred, Q , by passage of a

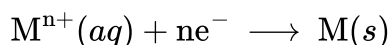
constant current, I , over a specified time interval, t , is then given by the simple mathematical product

Equation:

$$Q = It$$

When electrons are transferred during a redox process, the stoichiometry of the reaction may be used to derive the total amount of (electronic) charge involved. For example, the generic reduction process

Equation:



involves the transfer of n mole of electrons. The charge transferred is, therefore,

Equation:

$$Q = nF$$

where F is Faraday's constant, the charge in coulombs for one mole of electrons. If the reaction takes place in an electrochemical cell, the current flow is conveniently measured, and it may be used to assist in stoichiometric calculations related to the cell reaction.

Example:

Converting Current to Moles of Electrons

In one process used for electroplating silver, a current of 10.23 A was passed through an electrolytic cell for exactly 1 hour. How many moles of electrons passed through the cell? What mass of silver was deposited at the cathode from the silver nitrate solution?

Solution

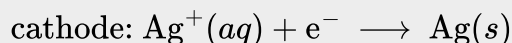
Faraday's constant can be used to convert the charge (Q) into moles of electrons (n). The charge is the current (I) multiplied by the time

Equation:

$$n = \frac{Q}{F} = \frac{\frac{10.23 \text{ C}}{\text{s}} \times 1 \text{ hr} \times \frac{60 \text{ min}}{\text{hr}} \times \frac{60 \text{ s}}{\text{min}}}{96,485 \text{ C/mol e}^{-}} = \frac{36,830 \text{ C}}{96,485 \text{ C/mol e}^{-}} = 0.3817 \text{ mol e}^{-}$$

From the problem, the solution contains AgNO_3 , so the reaction at the cathode involves 1 mole of electrons for each mole of silver

Equation:



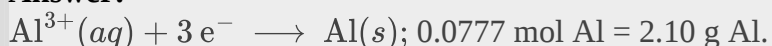
The atomic mass of silver is 107.9 g/mol, so

Equation:

$$\text{mass Ag} = 0.3817 \text{ mol e}^- \times \frac{1 \text{ mol Ag}}{1 \text{ mol e}^-} \times \frac{107.9 \text{ g Ag}}{1 \text{ mol Ag}} = 41.19 \text{ g Ag}$$

Check Your Learning

Aluminum metal can be made from aluminum(III) ions by electrolysis. What is the half-reaction at the cathode? What mass of aluminum metal would be recovered if a current of 25.0 A passed through the solution for 15.0 minutes?

Note:**Answer:****Example:****Time Required for Deposition**

In one application, a 0.010-mm layer of chromium must be deposited on a part with a total surface area of 3.3 m² from a solution of containing chromium(III) ions. How long would it take to deposit the layer of chromium if the current was 33.46 A? The density of chromium (metal) is 7.19 g/cm³.

Solution

First, compute the volume of chromium that must be produced (equal to the product of surface area and thickness):

Equation:

$$\text{volume} = \left(0.010 \text{ mm} \times \frac{1 \text{ cm}}{10 \text{ mm}} \right) \times \left(3.3 \text{ m}^2 \times \left(\frac{10,000 \text{ cm}^2}{1 \text{ m}^2} \right) \right) = 33 \text{ cm}^3$$

Use the computed volume and the provided density to calculate the molar amount of chromium required:

Equation:

$$\text{mass} = \text{volume} \times \text{density} = 33 \text{ cm}^3 \times \frac{7.19 \text{ g}}{\text{cm}^3} = 237 \text{ g Cr}$$

Equation:

$$\text{mol Cr} = 237 \text{ g Cr} \times \frac{1 \text{ mol Cr}}{52.00 \text{ g Cr}} = 4.56 \text{ mol Cr}$$

The stoichiometry of the chromium(III) reduction process requires three moles of electrons for each mole of chromium(0) produced, and so the total charge required is:

Equation:

$$Q = 4.56 \text{ mol Cr} \times \frac{3 \text{ mol e}^-}{1 \text{ mol Cr}} \times \frac{96485 \text{ C}}{\text{mol e}^-} = 1.32 \times 10^6 \text{ C}$$

Finally, if this charge is passed at a rate of 33.46 C/s, the required time is:

Equation:

$$t = \frac{Q}{I} = \frac{1.32 \times 10^6 \text{ C}}{33.46 \text{ C/s}} = 3.95 \times 10^4 \text{ s} = 11.0 \text{ hr}$$

Check Your Learning

What mass of zinc is required to galvanize the top of a $3.00 \text{ m} \times 5.50 \text{ m}$ sheet of iron to a thickness of 0.100 mm of zinc? If the zinc comes from a solution of $\text{Zn}(\text{NO}_3)_2$ and the current is 25.5 A, how long will it take to galvanize the top of the iron? The density of zinc is 7.140 g/cm^3 .

Note:

Answer:

11.8 kg Zn requires 382 hours.

Key Concepts and Summary

Nonspontaneous redox processes may be forced to occur in electrochemical cells by the application of an appropriate potential using an external power source—a process known as electrolysis. Electrolysis is the basis for certain ore refining processes, the industrial production of many chemical commodities, and the electroplating of metal coatings on various products. Measurement of the current flow during electrolysis permits stoichiometric calculations.

Key Equations

- $Q = I \times t = n \times F$

Chemistry End of Chapter Exercises

Exercise:

Problem:

If a 2.5 A current flows through a circuit for 35 minutes, how many coulombs of charge moved through the circuit?

Exercise:**Problem:**

For the scenario in the previous question, how many electrons moved through the circuit?

Exercise:**Problem:**

Write the half-reactions and cell reaction occurring during electrolysis of each molten salt below.

(a) CaCl_2

(b) LiH

(c) AlCl_3

(d) CrBr_3

Exercise:**Problem:**

What mass of each product is produced in each of the electrolytic cells of the previous problem if a total charge of $3.33 \times 10^5 \text{ C}$ passes through each cell?

Solution:

(a) mass Ca = 69.1 g; (b) mass Li = 23.9 g; (c) mass Al = 31.0 g;
mass Cl_2 = 122 g; mass H_2 = 3.48 g; mass Cl_2 = 122 g; (d)
mass Cr = 59.8 g
mass Br_2 = 276 g

Exercise:**Problem:**

How long would it take to reduce 1 mole of each of the following ions using the current indicated?

(a) Al^{3+} , 1.234 A

(b) Ca^{2+} , 22.2 A

(c) Cr^{5+} , 37.45 A

(d) Au^{3+} , 3.57 A

Exercise:

Problem:

A current of 2.345 A passes through the cell shown in [\[link\]](#) for 45 minutes. What is the volume of the hydrogen collected at room temperature if the pressure is exactly 1 atm? (Hint: Is hydrogen the only gas present above the water?)

Solution:

0.79 L

Exercise:

Problem:

An irregularly shaped metal part made from a particular alloy was galvanized with zinc using a $\text{Zn}(\text{NO}_3)_2$ solution. When a current of 2.599 A was used, it took exactly 1 hour to deposit a 0.01123-mm layer of zinc on the part. What was the total surface area of the part? The density of zinc is 7.140 g/cm^3 .

Glossary

electrolysis

process using electrical energy to cause a nonspontaneous process to occur

electrolytic cell

electrochemical cell in which an external source of electrical power is used to drive an otherwise nonspontaneous process

The Behavior of Light

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain the evidence for Maxwell’s electromagnetic model of light
- Describe the relationship between wavelength, frequency, and speed of light
- Discuss the particle model of light and the definition of photon
- Explain how and why the amount of light we see from an object depends upon its distance

Coded into the light and other kinds of radiation that reach us from objects in the universe is a wide range of information about what those objects are like and how they work. If we can decipher this code and read the messages it contains, we can learn an enormous amount about the cosmos without ever having to leave Earth or its immediate environment.

The visible light and other radiation we receive from the stars and planets is generated by processes at the atomic level—by changes in the way the parts of an atom interact and move. Thus, to appreciate how light is generated, we must explore how atoms work. There is a bit of irony in the fact that in order to understand some of the largest structures in the universe, we must become acquainted with some of the smallest.

Notice that we have twice used the phrase “light and other radiation.” One of the key ideas explored in this chapter is that visible light is not unique; it

is merely the most familiar example of a much larger family of radiation that can carry information to us.

The word “radiation” will be used frequently in this book, so it is important to understand what it means. In everyday language, “radiation” is often used to describe certain kinds of energetic subatomic particles released by radioactive materials in our environment. (An example is the kind of radiation used to treat some cancers.) But this is not what we mean when we use the word “radiation” in an astronomy text. *Radiation*, as used in this book, is a general term for waves (including light waves) that *radiate* outward from a source.

As we saw in [Orbits and Gravity](#), Newton’s theory of gravity accounts for the motions of planets as well as objects on Earth. Application of this theory to a variety of problems dominated the work of scientists for nearly two centuries. In the nineteenth century, many physicists turned to the study of electricity and magnetism, which are intimately connected with the production of light.

The scientist who played a role in this field comparable to Newton’s role in the study of gravity was physicist James Clerk Maxwell, born and educated in Scotland ([\[link\]](#)). Inspired by a number of ingenious experiments that showed an intimate relationship between electricity and magnetism, Maxwell developed a theory that describes both electricity and magnetism with only a small number of elegant equations. It is this theory that gives us important insights into the nature and behavior of light.
James Clerk Maxwell (1831–1879).



Maxwell unified the rules governing electricity and magnetism into a coherent theory.

Maxwell's Theory of Electromagnetism

We will look at the structure of the atom in more detail later, but we begin by noting that the typical atom consists of several types of particles, a number of which have not only mass but an additional property called electric charge. In the nucleus (central part) of every atom are *protons*, which are positively charged; outside the nucleus are electrons, which have a negative charge.

Maxwell's theory deals with these electric charges and their effects, especially when they are moving. In the vicinity of an electron charge, another charge feels a force of attraction or repulsion: opposite charges attract; like charges repel. When charges are not in motion, we observe only this electric attraction or repulsion. If charges are in motion, however (as

they are inside every atom and in a wire carrying a current), then we measure another force called *magnetism*.

Magnetism was well known for much of recorded human history, but its cause was not understood until the nineteenth century. Experiments with electric charges demonstrated that magnetism was the result of moving charged particles. Sometimes, the motion is clear, as in the coils of heavy wire that make an industrial electromagnet. Other times, it is more subtle, as in the kind of magnet you buy in a hardware store, in which many of the electrons inside the atoms are spinning in roughly the same direction; it is the alignment of their motion that causes the material to become magnetic.

Physicists use the word *field* to describe the action of forces that one object exerts on other distant objects. For example, we say the Sun produces a *gravitational field* that controls Earth's orbit, even though the Sun and Earth do not come directly into contact. Using this terminology, we can say that stationary electric charges produce *electric fields*, and moving electric charges also produce *magnetic fields*.

Actually, the relationship between electric and magnetic phenomena is even more profound. Experiments showed that changing magnetic fields could produce electric currents (and thus changing electric fields), and changing electric currents could in turn produce changing magnetic fields. So once begun, electric and magnetic field changes could continue to trigger each other.

Maxwell analyzed what would happen if electric charges were oscillating (moving constantly back and forth) and found that the resulting pattern of electric and magnetic fields would spread out and travel rapidly through space. Something similar happens when a raindrop strikes the surface of water or a frog jumps into a pond. The disturbance moves outward and creates a pattern we call a *wave* in the water ([\[link\]](#)). You might, at first, think that there must be very few situations in nature where electric charges oscillate, but this is not at all the case. As we shall see, atoms and molecules (which consist of charged particles) oscillate back and forth all the time. The resulting electromagnetic disturbances are among the most common phenomena in the universe.

Making Waves.



An oscillation in a pool of water creates an expanding disturbance called a wave. (credit: modification of work by "vastateparksstaff"/Flickr)

Maxwell was able to calculate the speed at which an electromagnetic disturbance moves through space; he found that it is equal to the speed of light, which had been measured experimentally. On that basis, he speculated that light was one form of a family of possible electromagnetic disturbances called **electromagnetic radiation**, a conclusion that was again confirmed in laboratory experiments. When light (reflected from the pages of an astronomy textbook, for example) enters a human eye, its changing electric and magnetic fields stimulate nerve endings, which then transmit the information contained in these changing fields to the brain. The science of astronomy is primarily about analyzing radiation from distant objects to understand what they are and how they work.

The Wave-Like Characteristics of Light

The changing electric and magnetic fields in light are similar to the waves that can be set up in a quiet pool of water. In both cases, the disturbance

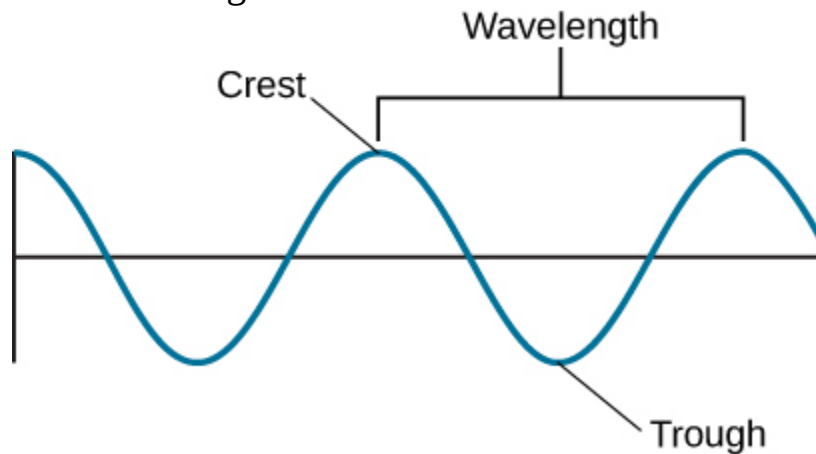
travels rapidly outward from the point of origin and can use its energy to disturb other things farther away. (For example, in water, the expanding ripples moving away from our frog could disturb the peace of a dragonfly resting on a leaf in the same pool.) In the case of electromagnetic waves, the radiation generated by a transmitting antenna full of charged particles and moving electrons at your local radio station can, sometime later, disturb a group of electrons in your car radio antenna and bring you the news and weather while you are driving to class or work in the morning.

The waves generated by charged particles differ from water waves in some profound ways, however. Water waves require water to travel in. The sound waves we hear, to give another example, are pressure disturbances that require air to travel through. But electromagnetic waves do not require water or air: the fields generate each other and so can move through a vacuum (such as outer space). This was such a disturbing idea to nineteenth-century scientists that they actually made up a substance to fill all of space—one for which there was not a single shred of evidence—just so light waves could have something to travel through: they called it the *aether*. Today, we know that there is no aether and that electromagnetic waves have no trouble at all moving through empty space (as all the starlight visible on a clear night must surely be doing).

The other difference is that *all* electromagnetic waves move at the same speed in empty space (the speed of light—approximately 300,000 kilometers per second, or 300,000,000 meters per second, which can also be written as 3×10^8 m/s), which turns out to be the fastest possible speed in the universe. No matter where electromagnetic waves are generated from and no matter what other properties they have, when they are moving (and not interacting with matter), they move at the speed of light. Yet you know from everyday experience that there are different kinds of light. For example, we perceive that light waves differ from one another in a property we call color. Let's see how we can denote the differences among the whole broad family of electromagnetic waves.

The nice thing about a wave is that it is a repeating phenomenon. Whether it is the up-and-down motion of a water wave or the changing electric and magnetic fields in a wave of light, the pattern of disturbance repeats in a

cyclical way. Thus, any wave motion can be characterized by a series of crests and troughs ([link](#)). Moving from one crest through a trough to the next crest completes one cycle. The horizontal length covered by one cycle is called the **wavelength**. Ocean waves provide an analogy: the wavelength is the distance that separates successive wave crests. Characterizing Waves.



Electromagnetic radiation has wave-like characteristics. The wavelength (λ) is the distance between crests, the frequency (f) is the number of cycles per second, and the speed (c) is the distance the wave covers during a specified period of time (e.g., kilometers per second).

For visible light, our eyes perceive different wavelengths as different colors: red, for example, is the longest visible wavelength, and violet is the shortest. The main colors of visible light from longest to shortest wavelength can be remembered using the mnemonic ROY G BIV—for Red, Orange, Yellow, Green, Blue, Indigo, and Violet. Other invisible forms of electromagnetic radiation have different wavelengths, as we will see in the next section.

We can also characterize different waves by their **frequency**, the number of wave cycles that pass by per second. If you count 10 crests moving by each

second, for example, then the frequency is 10 cycles per second (cps). In honor of Heinrich Hertz, the physicist who—inspired by Maxwell’s work—discovered radio waves, a cps is also called a *hertz* (Hz). Take a look at your radio, for example, and you will see the channel assigned to each radio station is characterized by its frequency, usually in units of KHz (kilohertz, or thousands of hertz) or MHz (megahertz, or millions of hertz).

Wavelength (λ) and frequency (f) are related because all electromagnetic waves travel at the same speed. To see how this works, imagine a parade in which everyone is forced by prevailing traffic conditions to move at exactly the same speed. You stand on a corner and watch the waves of marchers come by. First you see row after row of miniature ponies. Because they are not very large and, therefore, have a shorter wavelength, a good number of the ponies can move past you each minute; we can say they have a high frequency. Next, however, come several rows of circus elephants. The elephants are large and marching at the same speed as the ponies, so far fewer of them can march past you per minute: Because they have a wider spacing (longer wavelength), they represent a lower frequency.

The formula for this relationship can be expressed as follows: for any wave motion, the speed at which a wave moves equals the frequency times the wavelength. Waves with longer wavelengths have lower frequencies.

Mathematically, we can express this as

Equation:

$$c = \lambda f$$

where the Greek letter for “l”—lambda, λ —is used to denote wavelength and c is the scientific symbol for the speed of light. Solving for the wavelength, this is expressed as:

Equation:

$$\lambda = \frac{c}{f}.$$

Example:**Deriving and Using the Wave Equation**

The equation for the relationship between the speed and other characteristics of a wave can be derived from our basic understanding of motion. The average speed of anything that is moving is:

Equation:

$$\text{average speed} = \frac{\text{distance}}{\text{time}}$$

(So, for example, a car on the highway traveling at a speed of 100 km/h covers 100 km during the time of 1 h.) For an electromagnetic wave to travel the distance of one of its wavelengths, λ , at the speed of light, c , we have $c = \lambda/t$. The frequency of a wave is the number of cycles per second. If a wave has a frequency of a million cycles per second, then the time for each cycle to go by is a millionth of a second. So, in general, $t = 1/f$. Substituting into our wave equation, we get $c = \lambda \times f$. Now let's use this to calculate an example. What is the wavelength of visible light that has a frequency of 5.66×10^{14} Hz?

Solution

Solving the wave equation for wavelength, we find:

Equation:

$$\lambda = \frac{c}{f}$$

Substituting our values gives:

Equation:

$$\lambda = \frac{3.00 \times 10^8 \text{ m/s}}{5.66 \times 10^{14} \text{ Hz}} = 5.30 \times 10^{-7} \text{ m}$$

This answer can also be written as 530 nm, which is in the yellow-green part of the visible spectrum (nm stands for nanometers, where the term “nano” means “billionths”).

Check Your Learning

“Tidal waves,” or tsunamis, are waves caused by earthquakes that travel rapidly through the ocean. If a tsunami travels at the speed of 600 km/h

and approaches a shore at a rate of one wave crest every 15 min (4 waves/h), what would be the distance between those wave crests at sea?

Note:

Answer:

$$\lambda = \frac{600 \text{ km/h}}{4 \text{ waves/h}} = 150 \text{ km}$$

Light as a Photon

The electromagnetic wave model of light (as formulated by Maxwell) was one of the great triumphs of nineteenth-century science. In 1887, when Heinrich Hertz actually made invisible electromagnetic waves (what today are called radio waves) on one side of a room and detected them on the other side, it ushered in a new era that led to the modern age of telecommunications. His experiment ultimately led to the technologies of television, cell phones, and today's wireless networks around the globe.

However, by the beginning of the twentieth century, more sophisticated experiments had revealed that light behaves in certain ways that cannot be explained by the wave model. Reluctantly, physicists had to accept that sometimes light behaves more like a “particle”—or at least a self-contained packet of energy—than a wave. We call such a packet of electromagnetic energy a **photon**.

The fact that light behaves like a wave in certain experiments and like a particle in others was a very surprising and unlikely idea. After all, our common sense says that waves and particles are opposite concepts. On one hand, a wave is a repeating disturbance that, by its very nature, is not in only one place, but spreads out. A particle, on the other hand, is something that can be in only one place at any given time. Strange as it sounds,

though, countless experiments now confirm that electromagnetic radiation can sometimes behave like a wave and at other times like a particle.

Then, again, perhaps we shouldn't be surprised that something that always travels at the "speed limit" of the universe and doesn't need a medium to travel through might not obey our everyday common sense ideas. The confusion that this wave-particle duality of light caused in physics was eventually resolved by the introduction of a more complicated theory of waves and particles, now called quantum mechanics. (This is one of the most interesting fields of modern science, but it is mostly beyond the scope of our book. If you are interested in it, see some of the suggested resources at the end of this chapter.)

In any case, you should now be prepared when scientists (or the authors of this book) sometimes discuss electromagnetic radiation as if it consisted of waves and at other times refer to it as a stream of photons. A photon (being a packet of energy) carries a specific amount of energy. We can use the idea of energy to connect the photon and wave models. How much energy a photon has depends on its frequency when you think about it as a wave. A low-energy radio wave has a low frequency as a wave, while a high-energy X-ray at your dentist's office is a high-frequency wave. Among the colors of visible light, violet-light photons have the highest energy and red-light photons have the lowest.

Test whether the connection between photons and waves is clear to you. In the above example, which photon would have the longer wavelength as a wave: the radio wave or the X-ray? If you answered the radio wave, you are correct. Radio waves have a lower frequency, so the wave cycles are longer (they are elephants, not miniature ponies).

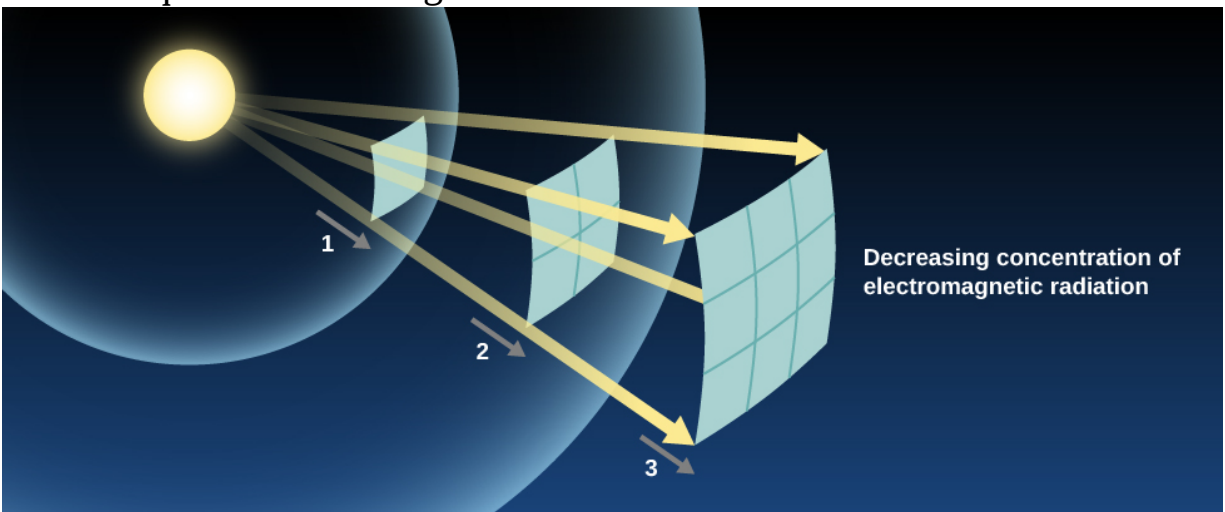
Propagation of Light

Let's think for a moment about how light from a lightbulb moves through space. As waves expand, they travel away from the bulb, not just toward your eyes but in all directions. They must therefore cover an ever-widening space. Yet the total amount of light available can't change once the light has left the bulb. This means that, as the same expanding shell of light covers a

larger and larger area, there must be less and less of it in any given place. Light (and all other electromagnetic radiation) gets weaker and weaker as it gets farther from its source.

The increase in the area that the light must cover is proportional to the square of the distance that the light has traveled ([link](#)). If we stand twice as far from the source, our eyes will intercept two-squared (2×2), or four times less light. If we stand 10 times farther from the source, we get 10-squared, or 100 times less light. You can see how this weakening means trouble for sources of light at astronomical distances. One of the nearest stars, Alpha Centauri A, emits about the same total energy as the Sun. But it is about 270,000 times farther away, and so it appears about 73 billion times fainter. No wonder the stars, which close-up would look more or less like the Sun, look like faint pinpoints of light from far away.

Inverse Square Law for Light.



As light radiates away from its source, it spreads out in such a way that the energy per unit area (the amount of energy passing through one of the small squares) decreases as the square of the distance from its source.

This idea—that the apparent brightness of a source (how bright it looks to us) gets weaker with distance in the way we have described—is known as the **inverse square law** for light propagation. In this respect, the

propagation of light is similar to the effects of gravity. Remember that the force of gravity between two attracting masses is also inversely proportional to the square of their separation.

Example:

The Inverse Square Law for Light

The intensity of a 120-W lightbulb observed from a distance 2 m away is 2.4 W/m^2 . What would be the intensity if this distance was doubled?

Solution

If we move twice as far away, then the answer will change according to the inverse square of the distance, so the new intensity will be $(1/2)^2 = 1/4$ of the original intensity, or 0.6 W/m^2 .

Check Your Learning

How many times brighter or fainter would a star appear if it were moved to:

- a. twice its present distance?
- b. ten times its present distance?
- c. half its present distance?

Note:

Answer:

a. $\left(\frac{1}{2}\right)^2 = \frac{1}{4}$; b. $\left(\frac{1}{10}\right)^2 = \frac{1}{100}$; c. $\left(\frac{1}{1/2}\right)^2 = 4$

Key Concepts and Summary

James Clerk Maxwell showed that whenever charged particles change their motion, as they do in every atom and molecule, they give off waves of

energy. Light is one form of this electromagnetic radiation. The wavelength of light determines the color of visible radiation. Wavelength (λ) is related to frequency (f) and the speed of light (c) by the equation $c = \lambda f$.

Electromagnetic radiation sometimes behaves like waves, but at other times, it behaves as if it were a particle—a little packet of energy, called a photon. The apparent brightness of a source of electromagnetic energy decreases with increasing distance from that source in proportion to the square of the distance—a relationship known as the inverse square law.

Glossary

electromagnetic radiation

radiation consisting of waves propagated through regularly varying electric and magnetic fields and traveling at the speed of light

frequency

the number of waves that cross a given point per unit time (in radiation)

inverse square law

(for light) the amount of energy (light) flowing through a given area in a given time decreases in proportion to the square of the distance from the source of energy or light

photon

a discrete unit (or “packet”) of electromagnetic energy

wavelength

the distance from crest to crest or trough to trough in a wave

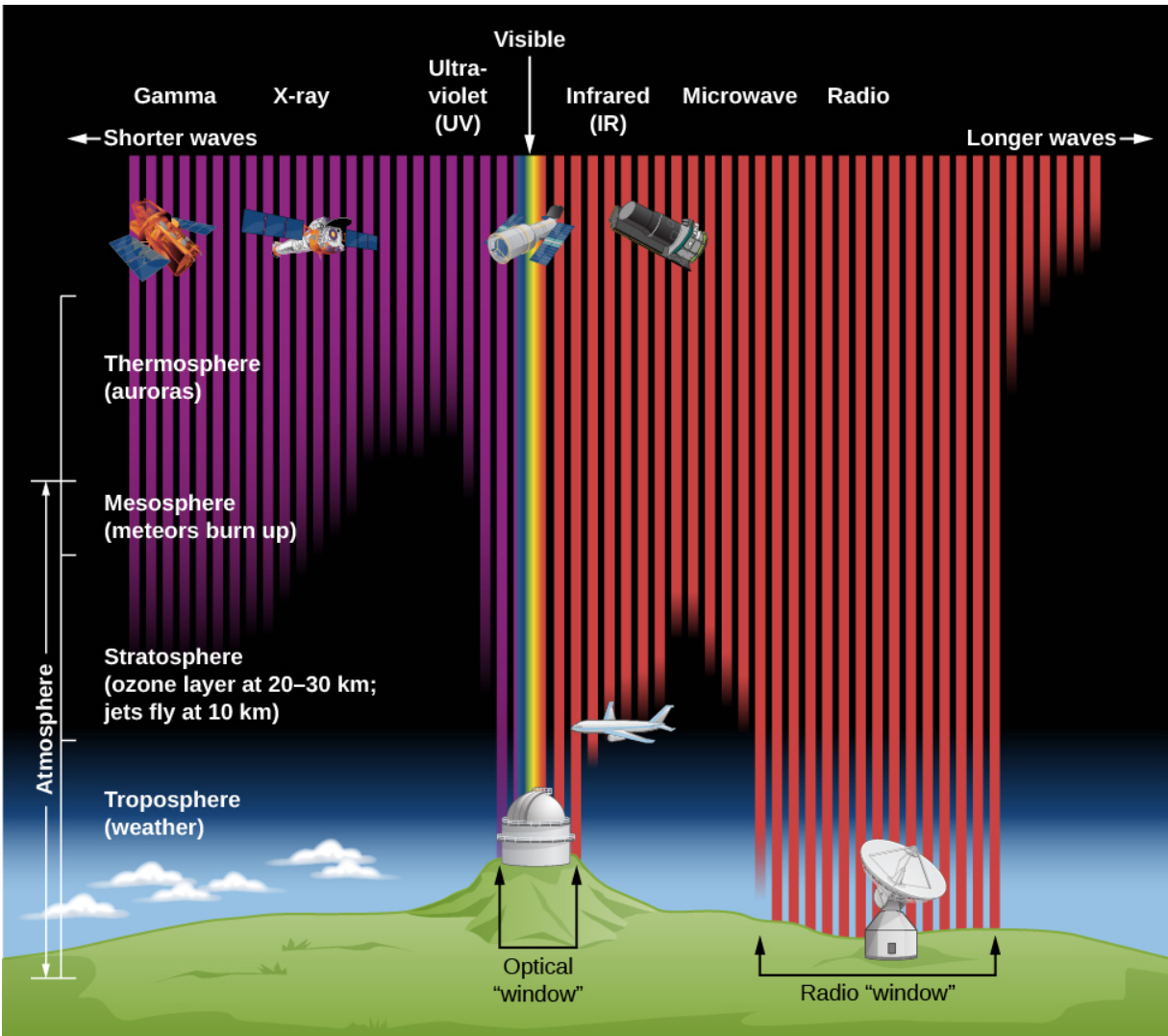
The Electromagnetic Spectrum

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Understand the bands of the electromagnetic spectrum and how they differ from one another
- Understand how each part of the spectrum interacts with Earth's atmosphere
- Explain how and why the light emitted by an object depends on its temperature

Objects in the universe send out an enormous range of electromagnetic radiation. Scientists call this range the **electromagnetic spectrum**, which they have divided into a number of categories. The spectrum is shown in [\[link\]](#), with some information about the waves in each part or band. Radiation and Earth's Atmosphere.



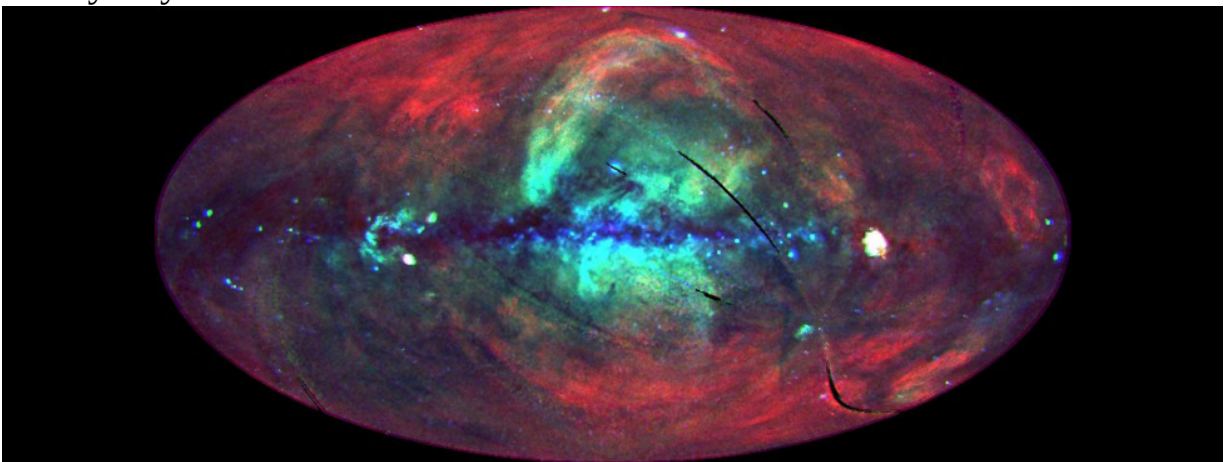
This figure shows the bands of the electromagnetic spectrum and how well Earth's atmosphere transmits them. Note that high-frequency waves from space do not make it to the surface and must therefore be observed from space. Some infrared and microwaves are absorbed by water and thus are best observed from high altitudes. Low-frequency radio waves are blocked by Earth's ionosphere. (credit: modification of work by STScI/JHU/NASA)

Types of Electromagnetic Radiation

Electromagnetic radiation with the shortest wavelengths, no longer than 0.01 nanometer, is categorized as **gamma rays** (1 nanometer = 10^{-9} meters; see [Appendix D](#)). The name *gamma* comes from the third letter of the Greek alphabet: gamma rays were the third kind of radiation discovered coming from radioactive atoms when physicists first investigated their behavior. Because gamma rays carry a lot of energy, they can be dangerous for living tissues. Gamma radiation is generated deep in the interior of stars, as well as by some of the most violent phenomena in the universe, such as the deaths of stars and the merging of stellar corpses. Gamma rays coming to Earth are absorbed by our atmosphere before they reach the ground (which is a good thing for our health); thus, they can only be studied using instruments in space.

Electromagnetic radiation with wavelengths between 0.01 nanometer and 20 nanometers is referred to as **X-rays**. Being more energetic than visible light, X-rays are able to penetrate soft tissues but not bones, and so allow us to make images of the shadows of the bones inside us. While X-rays can penetrate a short length of human flesh, they are stopped by the large numbers of atoms in Earth's atmosphere with which they interact. Thus, X-ray astronomy (like gamma-ray astronomy) could not develop until we invented ways of sending instruments above our atmosphere ([\[link\]](#)).

X-Ray Sky.



This is a map of the sky tuned to certain types of X-rays (seen from above Earth's atmosphere). The map tilts the sky so that the disk of our Milky Way Galaxy runs across its center. It was constructed and artificially colored from data gathered by the European ROSAT

satellite. Each color (red, yellow, and blue) shows X-rays of different frequencies or energies. For example, red outlines the glow from a hot local bubble of gas all around us, blown by one or more exploding stars in our cosmic vicinity. Yellow and blue show more distant sources of X-rays, such as remnants of other exploded stars or the active center of our Galaxy (in the middle of the picture). (credit: modification of work by NASA)

Radiation intermediate between X-rays and visible light is **ultraviolet** (meaning higher energy than violet). Outside the world of science, ultraviolet light is sometimes called “black light” because our eyes cannot see it. Ultraviolet radiation is mostly blocked by the ozone layer of Earth’s atmosphere, but a small fraction of ultraviolet rays from our Sun do penetrate to cause sunburn or, in extreme cases of overexposure, skin cancer in human beings. Ultraviolet astronomy is also best done from space.

Electromagnetic radiation with wavelengths between roughly 400 and 700 nm is called **visible light** because these are the waves that human vision can perceive. This is also the band of the electromagnetic spectrum that most readily reaches Earth’s surface. These two observations are not coincidental: human eyes evolved to see the kinds of waves that arrive from the Sun most effectively. Visible light penetrates Earth’s atmosphere effectively, except when it is temporarily blocked by clouds.

Between visible light and radio waves are the wavelengths of **infrared** or heat radiation. Astronomer William Herschel first discovered infrared in 1800 while trying to measure the temperatures of different colors of sunlight spread out into a spectrum. He noticed that when he accidentally positioned his thermometer beyond the reddest color, it still registered heating due to some invisible energy coming from the Sun. This was the first hint about the existence of the other (invisible) bands of the electromagnetic spectrum, although it would take many decades for our full understanding to develop.

A heat lamp radiates mostly infrared radiation, and the nerve endings in our skin are sensitive to this band of the electromagnetic spectrum. Infrared waves are absorbed by water and carbon dioxide molecules, which are more concentrated low in Earth's atmosphere. For this reason, infrared astronomy is best done from high mountaintops, high-flying airplanes, and spacecraft.

After infrared comes the familiar **microwave**, used in short-wave communication and microwave ovens. (Wavelengths vary from 1 millimeter to 1 meter and are absorbed by water vapor, which makes them effective in heating foods.) The “micro-” prefix refers to the fact that microwaves are small in comparison to radio waves, the next on the spectrum. You may remember that tea—which is full of water—heats up quickly in your microwave oven, while a ceramic cup—from which water has been removed by baking—stays cool in comparison.

All electromagnetic waves longer than microwaves are called **radio waves**, but this is so broad a category that we generally divide it into several subsections. Among the most familiar of these are radar waves, which are used in radar guns by traffic officers to determine vehicle speeds, and AM radio waves, which were the first to be developed for broadcasting. The wavelengths of these different categories range from over a meter to hundreds of meters, and other radio radiation can have wavelengths as long as several kilometers.

With such a wide range of wavelengths, not all radio waves interact with Earth's atmosphere in the same way. FM and TV waves are not absorbed and can travel easily through our atmosphere. AM radio waves are absorbed or reflected by a layer in Earth's atmosphere called the ionosphere (the ionosphere is a layer of charged particles at the top of our atmosphere, produced by interactions with sunlight and charged particles that are ejected from the Sun).

We hope this brief survey has left you with one strong impression: although visible light is what most people associate with astronomy, the light that our eyes can see is only a tiny fraction of the broad range of waves generated in the universe. Today, we understand that judging some astronomical phenomenon by using only the light we can see is like hiding under the table at a big dinner party and judging all the guests by nothing but their

shoes. There’s a lot more to each person than meets our eye under the table. It is very important for those who study astronomy today to avoid being “visible light chauvinists”—to respect only the information seen by their eyes while ignoring the information gathered by instruments sensitive to other bands of the electromagnetic spectrum.

[\[link\]](#) summarizes the bands of the electromagnetic spectrum and indicates the temperatures and typical astronomical objects that emit each kind of electromagnetic radiation. While at first, some of the types of radiation listed in the table may seem unfamiliar, you will get to know them better as your astronomy course continues. You can return to this table as you learn more about the types of objects astronomers study.

Types of Electromagnetic Radiation			
Type of Radiation	Wavelength Range (nm)	Radiated by Objects at This Temperature	Typical Sources
Gamma rays	Less than 0.01	More than 10^8 K	Produced in nuclear reactions; require very high-energy processes
X-rays	0.01–20	10^6 – 10^8 K	Gas in clusters of galaxies, supernova remnants, solar corona

Types of Electromagnetic Radiation			
Type of Radiation	Wavelength Range (nm)	Radiated by Objects at This Temperature	Typical Sources
Ultraviolet	20–400	10^4 – 10^6 K	Supernova remnants, very hot stars
Visible	400–700	10^3 – 10^4 K	Stars
Infrared	10^3 – 10^6	10 – 10^3 K	Cool clouds of dust and gas, planets, moons
Microwave	10^6 – 10^9	Less than 10 K	Active galaxies, pulsars, cosmic background radiation
Radio	More than 10^9	Less than 10 K	Supernova remnants, pulsars, cold gas

Radiation and Temperature

Some astronomical objects emit mostly infrared radiation, others mostly visible light, and still others mostly ultraviolet radiation. What determines the type of electromagnetic radiation emitted by the Sun, stars, and other dense astronomical objects? The answer often turns out to be their *temperature*.

At the microscopic level, everything in nature is in motion. A solid is composed of molecules and atoms in continuous vibration: they move back and forth in place, but their motion is much too small for our eyes to make out. A gas consists of atoms and/or molecules that are flying about freely at high speed, continually bumping into one another and bombarding the surrounding matter. The hotter the solid or gas, the more rapid the motion of its molecules or atoms. The temperature of something is thus a measure of the average motion energy of the particles that make it up.

This motion at the microscopic level is responsible for much of the electromagnetic radiation on Earth and in the universe. As atoms and molecules move about and collide, or vibrate in place, their electrons give off electromagnetic radiation. The characteristics of this radiation are determined by the temperature of those atoms and molecules. In a hot material, for example, the individual particles vibrate in place or move rapidly from collisions, so the emitted waves are, on average, more energetic. And recall that higher energy waves have a higher frequency. In very cool material, the particles have low-energy atomic and molecular motions and thus generate lower-energy waves.

Note:

Check out the [NASA briefing](#) or [NASA's 5-minute introductory video](#) to learn more about the electromagnetic spectrum.

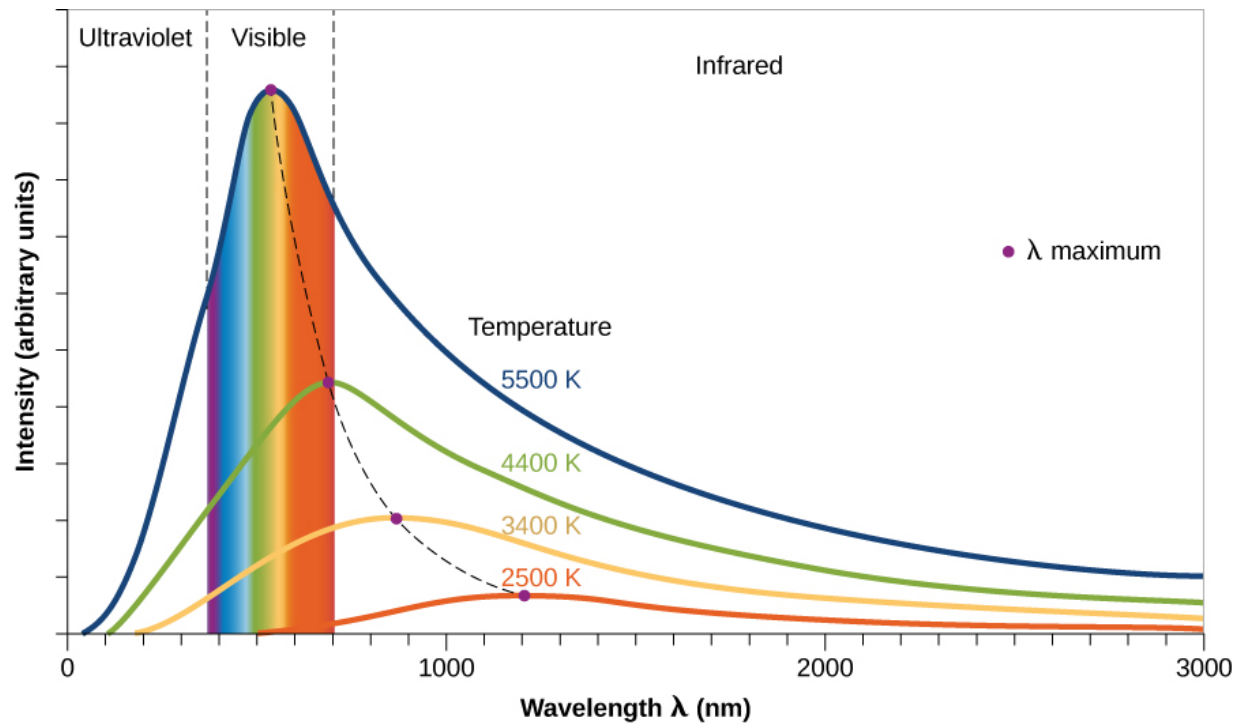
Radiation Laws

To understand, in more quantitative detail, the relationship between temperature and electromagnetic radiation, we imagine an idealized object called a **blackbody**. Such an object (unlike your sweater or your astronomy instructor's head) does not reflect or scatter any radiation, but absorbs all the electromagnetic energy that falls onto it. The energy that is absorbed causes the atoms and molecules in it to vibrate or move around at increasing speeds. As it gets hotter, this object will radiate electromagnetic waves until absorption and radiation are in balance. We want to discuss such an

idealized object because, as you will see, stars behave in very nearly the same way.

The radiation from a blackbody has several characteristics, as illustrated in [\[link\]](#). The graph shows the power emitted at each wavelength by objects of different temperatures. In science, the word *power* means the energy coming off per second (and it is typically measured in *watts*, which you are probably familiar with from buying lightbulbs).

Radiation Laws Illustrated.



This graph shows in arbitrary units how many photons are given off at each wavelength for objects at four different temperatures. The wavelengths corresponding to visible light are shown by the colored bands. Note that at hotter temperatures, more energy (in the form of photons) is emitted at all wavelengths. The higher the temperature, the shorter the wavelength at which the peak amount of energy is radiated (this is known as Wien's law).

First of all, notice that the curves show that, at each temperature, our blackbody object emits radiation (photons) at all wavelengths (all colors).

This is because in any solid or denser gas, some molecules or atoms vibrate or move between collisions slower than average and some move faster than average. So when we look at the electromagnetic waves emitted, we find a broad range, or spectrum, of energies and wavelengths. More energy is emitted at the average vibration or motion rate (the highest part of each curve), but if we have a large number of atoms or molecules, some energy will be detected at each wavelength.

Second, note that an object at a higher temperature emits more power at all wavelengths than does a cooler one. In a hot gas (the taller curves in [\[link\]](#)), for example, the atoms have more collisions and give off more energy. In the real world of stars, this means that hotter stars give off more energy at every wavelength than do cooler stars.

Third, the graph shows us that the higher the temperature, the shorter the wavelength at which the maximum power is emitted. Remember that a shorter wavelength means a higher frequency and energy. It makes sense, then, that hot objects give off a larger fraction of their energy at shorter wavelengths (higher energies) than do cool objects. You may have observed examples of this rule in everyday life. When a burner on an electric stove is turned on low, it emits only heat, which is infrared radiation, but does not glow with visible light. If the burner is set to a higher temperature, it starts to glow a dull red. At a still-higher setting, it glows a brighter orange-red (shorter wavelength). At even higher temperatures, which cannot be reached with ordinary stoves, metal can appear brilliant yellow or even blue-white.

We can use these ideas to come up with a rough sort of “thermometer” for measuring the temperatures of stars. Because many stars give off most of their energy in visible light, the color of light that dominates a star’s appearance is a rough indicator of its temperature. If one star looks red and another looks blue, which one has the higher temperature? Because blue is the shorter-wavelength color, it is the sign of a hotter star. (Note that the temperatures we associate with different colors in science are not the same as the ones artists use. In art, red is often called a “hot” color and blue a “cool” color. Likewise, we commonly see red on faucet or air conditioning controls to indicate hot temperatures and blue to indicate cold temperatures.

Although these are common uses to us in daily life, in nature, it's the other way around.)

We can develop a more precise star thermometer by measuring how much energy a star gives off at each wavelength and by constructing diagrams like [\[link\]](#). The location of the peak (or maximum) in the power curve of each star can tell us its temperature. The average temperature at the surface of the Sun, which is where the radiation that we see is emitted, turns out to be 5800 K. (Throughout this text, we use the kelvin or absolute temperature scale. On this scale, water freezes at 273 K and boils at 373 K. All molecular motion ceases at 0 K. The various temperature scales are described in [Appendix D](#).) There are stars cooler than the Sun and stars hotter than the Sun.

The wavelength at which maximum power is emitted can be calculated according to the equation

Equation:

$$\lambda_{\text{max}} = \frac{3 \times 10^6}{T}$$

where the wavelength is in nanometers (one billionth of a meter) and the temperature is in K (the constant 3×10^6 has units of $\text{nm} \times \text{K}$). This relationship is called **Wien's law**. For the Sun, the wavelength at which the maximum energy is emitted is 520 nanometers, which is near the middle of that portion of the electromagnetic spectrum called visible light.

Characteristic temperatures of other astronomical objects, and the wavelengths at which they emit most of their power, are listed in [\[link\]](#).

Example:

Calculating the Temperature of a Blackbody

We can use Wien's law to calculate the temperature of a star provided we know the wavelength of peak intensity for its spectrum. If the emitted radiation from a red dwarf star has a wavelength of maximum power at 1200 nm, what is the temperature of this star, assuming it is a blackbody?

Solution

Solving Wien's law for temperature gives:

Equation:

$$T = \frac{3 \times 10^6 \text{ nm K}}{\lambda_{\text{max}}} = \frac{3 \times 10^6 \text{ nm K}}{1200 \text{ nm}} = 2500 \text{ K}$$

Check Your Learning

What is the temperature of a star whose maximum light is emitted at a much shorter wavelength of 290 nm?

Note:**Answer:**

$$T = \frac{3 \times 10^6 \text{ nm K}}{\lambda_{\text{max}}} = \frac{3 \times 10^6 \text{ nm K}}{290 \text{ nm}} = 10,300 \text{ K}$$

Since this star has a peak wavelength that is at a shorter wavelength (in the ultraviolet part of the spectrum) than that of our Sun (in the visible part of the spectrum), it should come as no surprise that its surface temperature is much hotter than our Sun's.

We can also describe our observation that hotter objects radiate more power at all wavelengths in a mathematical form. If we sum up the contributions from all parts of the electromagnetic spectrum, we obtain the total energy emitted by a blackbody. What we usually measure from a large object like a star is the **energy flux**, the power emitted per square meter. The word *flux* means “flow” here: we are interested in the flow of power into an area (like the area of a telescope mirror). It turns out that the energy flux from a blackbody at temperature T is proportional to the fourth power of its absolute temperature. This relationship is known as the **Stefan-Boltzmann law** and can be written in the form of an equation as

Equation:

$$F = \sigma T^4$$

where F stands for the energy flux (in units of watts per square meter), T is given in Kelvins, and σ (Greek letter sigma) is a constant number (5.67×10^{-8}).

Notice how impressive this result is. Increasing the temperature of a star would have a tremendous effect on the power it radiates. If the Sun, for example, were twice as hot—that is, if it had a temperature of 11,600 K—it would radiate 2^4 , or 16 times more power than it does now. Tripling the temperature would raise the power output 81 times. Hot stars really shine away a tremendous amount of energy.

Example:**Calculating the Power of a Star**

While energy flux tells us how much power a star emits per square meter, we would often like to know how much total power is emitted by the star. We can determine that by multiplying the energy flux by the number of square meters on the surface of the star. Stars are mostly spherical, so we can use the formula $4\pi R^2$ for the surface area, where R is the radius of the star. The total power emitted by the star (which we call the star's “absolute luminosity”) can be found by multiplying the formula for energy flux and the formula for the surface area:

Equation:

$$L = 4\pi R^2 \sigma T^4$$

Two stars have the same size and are the same distance from us. Star A has a surface temperature of 6000 K, and star B has a surface temperature twice as high, 12,000 K. How much more luminous is star B compared to star A?

Solution**Equation:**

$$L_A = 4\pi R_A^2 \sigma T_A^4 \text{ and } L_B = 4\pi R_B^2 \sigma T_B^4$$

Take the ratio of the luminosity of Star A to Star B:

Equation:

$$\frac{L_B}{L_A} = \frac{4\pi R_B^2 \sigma T_B^4}{4\pi R_A^2 \sigma T_A^4} = \frac{R_B^2 T_B^4}{R_A^2 T_A^4}$$

Because the two stars are the same size, $R_A = R_B$, leaving

Equation:

$$\frac{T_B^4}{T_A^4} = \frac{(12,000 \text{ K})^4}{(6000 \text{ K})^4} = 2^4 = 16$$

Check Your Learning

Two stars with identical diameters are the same distance away. One has a temperature of 8700 K and the other has a temperature of 2900 K. Which is brighter? How much brighter is it?

Note:

Answer:

The 8700 K star has triple the temperature, so it is $3^4 = 81$ times brighter.

Key Concepts and Summary

The electromagnetic spectrum consists of gamma rays, X-rays, ultraviolet radiation, visible light, infrared, and radio radiation. Many of these wavelengths cannot penetrate the layers of Earth's atmosphere and must be observed from space, whereas others—such as visible light, FM radio and TV—can penetrate to Earth's surface. The emission of electromagnetic

radiation is intimately connected to the temperature of the source. The higher the temperature of an idealized emitter of electromagnetic radiation, the shorter is the wavelength at which the maximum amount of radiation is emitted. The mathematical equation describing this relationship is known as Wien's law: $\lambda_{\text{max}} = (3 \times 10^6)/T$. The total power emitted per square meter increases with increasing temperature. The relationship between emitted energy flux and temperature is known as the Stefan-Boltzmann law: $F = \sigma T^4$.

Glossary

blackbody

an idealized object that absorbs all electromagnetic energy that falls onto it

electromagnetic spectrum

the whole array or family of electromagnetic waves, from radio to gamma rays

energy flux

the amount of energy passing through a unit area (for example, 1 square meter) per second; the units of flux are watts per square meter

gamma rays

photons (of electromagnetic radiation) of energy with wavelengths no longer than 0.01 nanometer; the most energetic form of electromagnetic radiation

infrared

electromagnetic radiation of wavelength 10^3 – 10^6 nanometers; longer than the longest (red) wavelengths that can be perceived by the eye, but shorter than radio wavelengths

microwave

electromagnetic radiation of wavelengths from 1 millimeter to 1 meter; longer than infrared but shorter than radio waves

radio waves

all electromagnetic waves longer than microwaves, including radar waves and AM radio waves

Stefan-Boltzmann law

a formula from which the rate at which a blackbody radiates energy can be computed; the total rate of energy emission from a unit area of a blackbody is proportional to the fourth power of its absolute temperature: $F = \sigma T^4$

ultraviolet

electromagnetic radiation of wavelengths 10 to 400 nanometers; shorter than the shortest visible wavelengths

visible light

electromagnetic radiation with wavelengths of roughly 400–700 nanometers; visible to the human eye

Wien's law

formula that relates the temperature of a blackbody to the wavelength at which it emits the greatest intensity of radiation

X-rays

electromagnetic radiation with wavelengths between 0.01 nanometer and 20 nanometers; intermediate between those of ultraviolet radiation and gamma rays

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the properties of light
- Explain how astronomers learn the composition of a gas by examining its spectral lines
- Discuss the various types of spectra

Electromagnetic radiation carries a lot of information about the nature of stars and other astronomical objects. To extract this information, however, astronomers must be able to study the amounts of energy we receive at different wavelengths of light in fine detail. Let's examine how we can do this and what we can learn.

Properties of Light

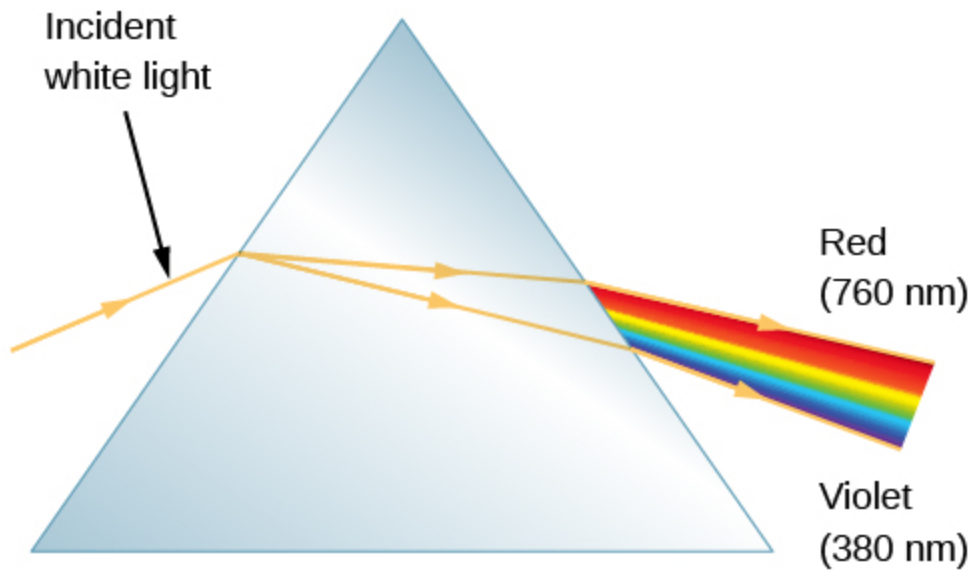
Light exhibits certain behaviors that are important to the design of telescopes and other instruments. For example, light can be *reflected* from a surface. If the surface is smooth and shiny, as with a mirror, the direction of the reflected light beam can be calculated accurately from knowledge of the shape of the reflecting surface. Light is also bent, or *refracted*, when it passes from one kind of transparent material into another—say, from the air into a glass lens.

Reflection and refraction of light are the basic properties that make possible all *optical* instruments (devices that help us to see things better)—from

eyeglasses to giant astronomical telescopes. Such instruments are generally combinations of glass lenses, which bend light according to the principles of refraction, and curved mirrors, which depend on the properties of reflection. Small optical devices, such as eyeglasses or binoculars, generally use lenses, whereas large telescopes depend almost entirely on mirrors for their main optical elements. We will discuss astronomical instruments and their uses more fully in [Astronomical Instruments](#). For now, we turn to another behavior of light, one that is essential for the decoding of light.

In 1672, in the first paper that he submitted to the Royal Society, Sir Isaac Newton described an experiment in which he permitted sunlight to pass through a small hole and then through a prism. Newton found that sunlight, which looks white to us, is actually made up of a mixture of all the colors of the rainbow ([link](#)).

Action of a Prism.



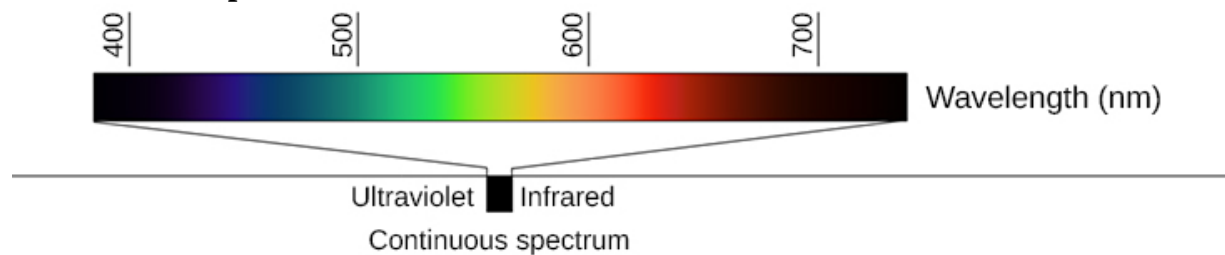
When we pass a beam of white sunlight through a prism, we see a rainbow-colored band of light that we call a continuous spectrum.

[link](#) shows how light is separated into different colors with a prism—a piece of glass in the shape of a triangle with refracting surfaces. Upon entering one face of the prism, the path of the light is refracted (bent), but

not all of the colors are bent by the same amount. The bending of the beam depends on the wavelength of the light as well as the properties of the material, and as a result, different wavelengths (or colors of light) are bent by different amounts and therefore follow slightly different paths through the prism. The violet light is bent more than the red. This phenomenon is called **dispersion** and explains Newton's rainbow experiment.

Upon leaving the opposite face of the prism, the light is bent again and further dispersed. If the light leaving the prism is focused on a screen, the different wavelengths or colors that make up white light are lined up side by side just like a rainbow ([\[link\]](#)). (In fact, a rainbow is formed by the dispersion of light through raindrops; see [The Rainbow](#) feature box.) Because this array of colors is a spectrum of light, the instrument used to disperse the light and form the spectrum is called a **spectrometer**.

Continuous Spectrum.



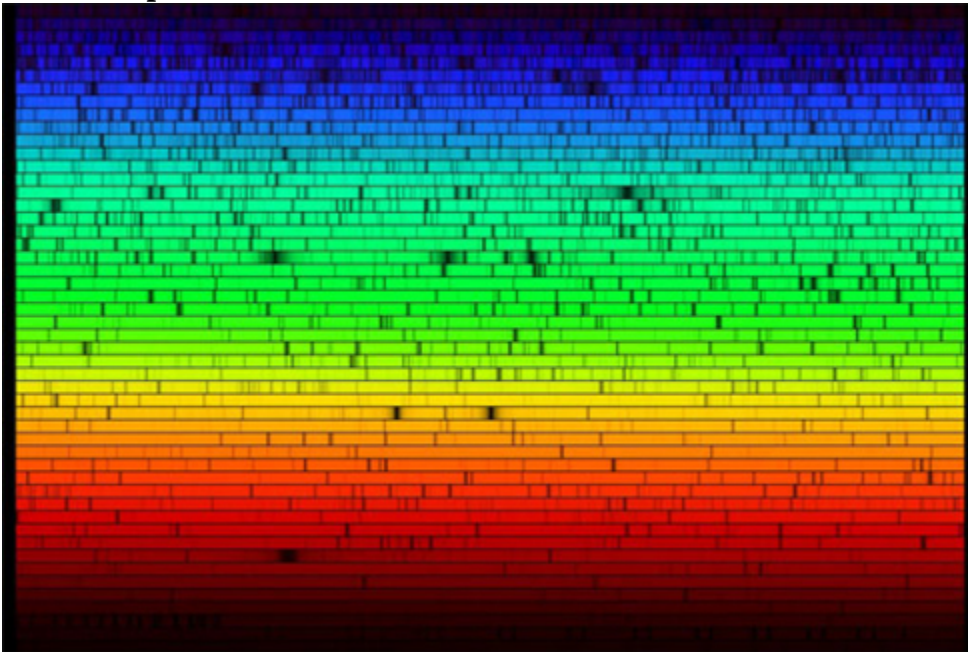
When white light passes through a prism, it is dispersed and forms a continuous spectrum of all the colors. Although it is hard to see in this printed version, in a well-dispersed spectrum, many subtle gradations in color are visible as your eye scans from one end (violet) to the other (red).

The Value of Stellar Spectra

When Newton described the laws of refraction and dispersion in optics, and observed the solar spectrum, all he could see was a continuous band of colors. If the spectrum of the white light from the Sun and stars were simply a continuous rainbow of colors, astronomers would have little interest in the detailed study of a star's spectrum once they had learned its average surface

temperature. In 1802, however, William Wollaston built an improved spectrometer that included a lens to focus the Sun's spectrum on a screen. With this device, Wollaston saw that the colors were not spread out uniformly, but instead, some ranges of color were missing, appearing as dark bands in the solar spectrum. He mistakenly attributed these lines to natural boundaries between the colors. In 1815, German physicist Joseph Fraunhofer, upon a more careful examination of the solar spectrum, found about 600 such dark lines (missing colors), which led scientists to rule out the boundary hypothesis ([\[link\]](#)).

Visible Spectrum of the Sun.



Our star's spectrum is crossed by dark lines produced by atoms in the solar atmosphere that absorb light at certain wavelengths. (credit: modification of work by Nigel Sharp, NOAO/National Solar Observatory at Kitt Peak/AURA, and the National Science Foundation)

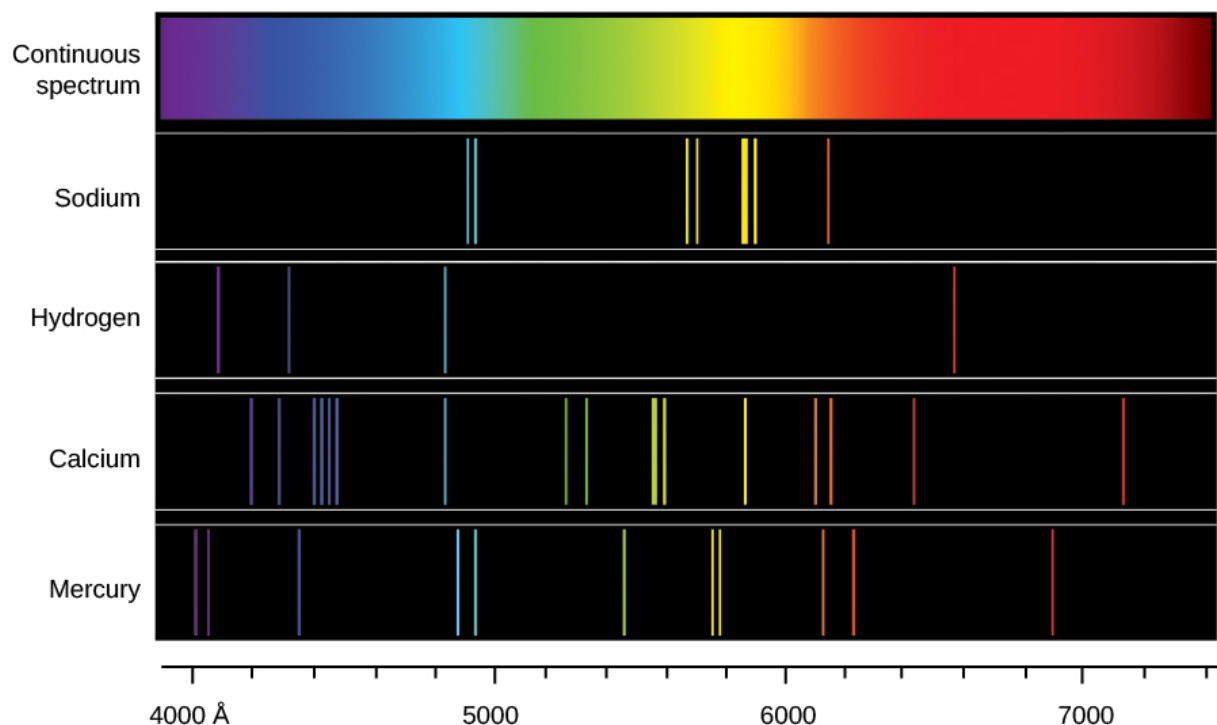
Later, researchers found that similar dark lines could be produced in the spectra ("spectra" is the plural of "spectrum") of artificial light sources. They did this by passing their light through various apparently transparent substances—usually containers with just a bit of thin gas in them.

These gases turned out not to be transparent at *all* colors: they were quite opaque at a few sharply defined wavelengths. Something in each gas had to be absorbing just a few colors of light and no others. All gases did this, but each different element absorbed a different set of colors and thus showed different dark lines. If the gas in a container consisted of two elements, then light passing through it was missing the colors (showing dark lines) for both of the elements. So it became clear that certain lines in the spectrum “go with” certain elements. This discovery was one of the most important steps forward in the history of astronomy.

What would happen if there were no continuous spectrum for our gases to remove light from? What if, instead, we heated the same thin gases until they were hot enough to glow with their own light? When the gases were heated, a spectrometer revealed no continuous spectrum, but several separate bright lines. That is, these hot gases emitted light only at certain specific wavelengths or colors.

When the gas was pure hydrogen, it would emit one pattern of colors; when it was pure sodium, it would emit a different pattern. A mixture of hydrogen and sodium emitted both sets of spectral lines. The colors the gases emitted when they were heated were the very same colors as those they had absorbed when a continuous source of light was behind them. From such experiments, scientists began to see that different substances showed distinctive *spectral signatures* by which their presence could be detected ([\[link\]](#)). Just as your signature allows the bank to identify you, the unique pattern of colors for each type of atom (its spectrum) can help us identify which element or elements are in a gas.

Continuous Spectrum and Line Spectra from Different Elements.



Each type of glowing gas (each element) produces its own unique pattern of lines, so the composition of a gas can be identified by its spectrum. The spectra of sodium, hydrogen, calcium, and mercury gases are shown here.

Types of Spectra

In these experiments, then, there were three different types of spectra. A **continuous spectrum** (formed when a solid or very dense gas gives off radiation) is an array of all wavelengths or colors of the rainbow. A continuous spectrum can serve as a backdrop from which the atoms of much less dense gas can absorb light. A dark line, or **absorption spectrum**, consists of a series or pattern of dark lines—missing colors—superimposed upon the continuous spectrum of a source. A bright line, or **emission spectrum**, appears as a pattern or series of bright lines; it consists of light in which only certain discrete wavelengths are present. ([link](#)) shows an absorption spectrum, whereas [link](#) shows the emission spectrum of a

number of common elements along with an example of a continuous spectrum.)

When we have a hot, thin gas, each particular chemical element or compound produces its own characteristic pattern of spectral lines—its spectral signature. No two types of atoms or molecules give the same patterns. In other words, each particular gas can absorb or emit only certain wavelengths of the light peculiar to that gas. In contrast, absorption spectra occur when passing white light through a cool, thin gas. The temperature and other conditions determine whether the lines are bright or dark (whether light is absorbed or emitted), but the wavelengths of the lines for any element are the same in either case. It is the precise pattern of wavelengths that makes the signature of each element unique. Liquids and solids can also generate spectral lines or bands, but they are broader and less well defined—and hence, more difficult to interpret. Spectral analysis, however, can be quite useful. It can, for example, be applied to light reflected off the surface of a nearby asteroid as well as to light from a distant galaxy.

The dark lines in the solar spectrum thus give evidence of certain chemical elements between us and the Sun absorbing those wavelengths of sunlight. Because the space between us and the Sun is pretty empty, astronomers realized that the atoms doing the absorbing must be in a thin atmosphere of cooler gas around the Sun. This outer atmosphere is not all that different from the rest of the Sun, just thinner and cooler. Thus, we can use what we learn about its composition as an indicator of what the whole Sun is made of. Similarly, we can use the presence of absorption and emission lines to analyze the composition of other stars and clouds of gas in space.

Such analysis of spectra is the key to modern astronomy. Only in this way can we “sample” the stars, which are too far away for us to visit. Encoded in the electromagnetic radiation from celestial objects is clear information about the chemical makeup of these objects. Only by understanding what the stars were made of could astronomers begin to form theories about what made them shine and how they evolved.

In 1860, German physicist Gustav Kirchhoff became the first person to use spectroscopy to identify an element in the Sun when he found the spectral signature of sodium gas. In the years that followed, astronomers found

many other chemical elements in the Sun and stars. In fact, the element helium was found first in the Sun from its spectrum and only later identified on Earth. (The word “helium” comes from *helios*, the Greek name for the Sun.)

Why are there specific lines for each element? The answer to that question was not found until the twentieth century; it required the development of a model for the atom. We therefore turn next to a closer examination of the atoms that make up all matter.

Note:

The Rainbow

Rainbows are an excellent illustration of the dispersion of sunlight. You have a good chance of seeing a rainbow any time you are between the Sun and a rain shower, as illustrated in [\[link\]](#). The raindrops act like little prisms and break white light into the spectrum of colors. Suppose a ray of sunlight encounters a raindrop and passes into it. The light changes direction—is refracted—when it passes from air to water; the blue and violet light are refracted more than the red. Some of the light is then reflected at the backside of the drop and reemerges from the front, where it is again refracted. As a result, the white light is spread out into a rainbow of colors.

Rainbow Refraction.



(a) This diagram shows how light from the Sun, which is located behind the observer, can be refracted by raindrops to produce (b) a rainbow. (c) Refraction separates white light into its component colors.

Note that violet light lies above the red light after it emerges from the raindrop. When you look at a rainbow, however, the red light is higher in the sky. Why? Look again at [\[link\]](#). If the observer looks at a raindrop that is high in the sky, the violet light passes over her head and the red light enters her eye. Similarly, if the observer looks at a raindrop that is low in the sky, the violet light reaches her eye and the drop appears violet, whereas the red light from that same drop strikes the ground and is not seen. Colors of intermediate wavelengths are refracted to the eye by drops that are intermediate in altitude between the drops that appear violet and the ones that appear red. Thus, a single rainbow always has red on the outside and violet on the inside.

Key Concepts and Summary

A spectrometer is a device that forms a spectrum, often utilizing the phenomenon of dispersion. The light from an astronomical source can consist of a continuous spectrum, an emission (bright line) spectrum, or an absorption (dark line) spectrum. Because each element leaves its spectral signature in the pattern of lines we observe, spectral analyses reveal the composition of the Sun and stars.

Glossary

absorption spectrum

a series or pattern of dark lines superimposed on a continuous spectrum

continuous spectrum

a spectrum of light composed of radiation of a continuous range of wavelengths or colors, rather than only certain discrete wavelengths

dispersion

separation of different wavelengths of white light through refraction of different amounts

emission spectrum

a series or pattern of bright lines superimposed on a continuous spectrum

spectrometer

an instrument for obtaining a spectrum; in astronomy, usually attached to a telescope to record the spectrum of a star, galaxy, or other astronomical object

The Structure of the Atom

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the structure of atoms and the components of nuclei
- Explain the behavior of electrons within atoms and how electrons interact with light to move among energy levels

The idea that matter is composed of tiny particles called atoms is at least 25 centuries old. It took until the twentieth century, however, for scientists to invent instruments that permitted them to probe inside an atom and find that it is not, as had been thought, hard and indivisible. Instead, the atom is a complex structure composed of still smaller particles.

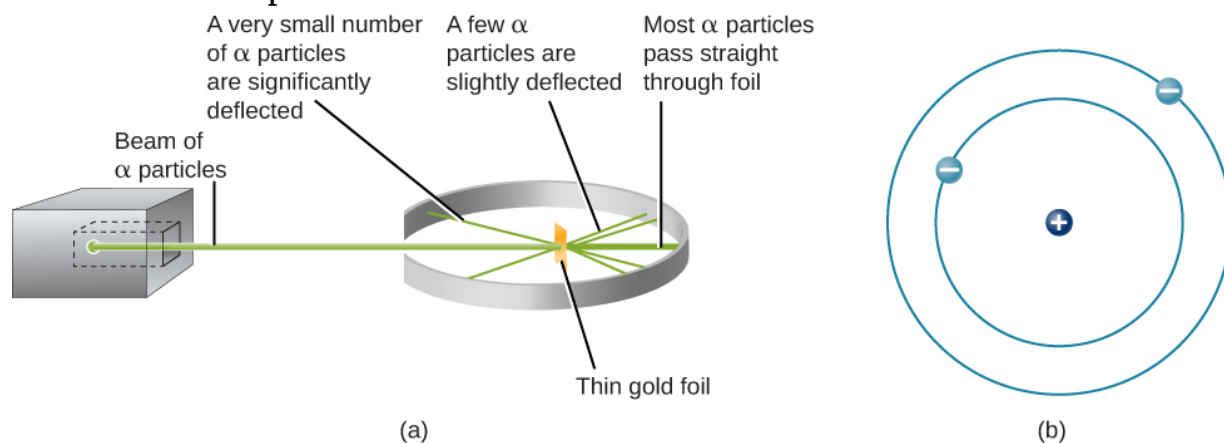
Probing the Atom

The first of these smaller particles was discovered by British physicist James (J. J.) Thomson in 1897. Named the *electron*, this particle is negatively charged. (It is the flow of these particles that produces currents of electricity, whether in lightning bolts or in the wires leading to your lamp.) Because an atom in its normal state is electrically neutral, each electron in an atom must be balanced by the same amount of positive charge.

The next step was to determine where in the atom the positive and negative charges are located. In 1911, British physicist Ernest Rutherford devised an experiment that provided part of the answer to this question. He bombarded

an extremely thin piece of gold foil, only about 400 atoms thick, with a beam of alpha particles ([link](#)). *Alpha particles* (α particles) are helium atoms that have lost their electrons and thus are positively charged. Most of these particles passed through the gold foil just as if it and the atoms in it were nearly empty space. About 1 in 8000 of the alpha particles, however, completely reversed direction and bounced backward from the foil. Rutherford wrote, “It was quite the most incredible event that has ever happened to me in my life. It was almost as incredible as if you fired a 15-inch shell at a piece of tissue paper and it came back and hit you.”

Rutherford's Experiment.



(a) When Rutherford allowed α particles from a radioactive source to strike a target of gold foil, he found that, although most of them went straight through, some rebounded back in the direction from which they came. (b) From this experiment, he concluded that the atom must be constructed like a miniature solar system, with the positive charge concentrated in the nucleus and the negative charge orbiting in the large volume around the nucleus. Note that this drawing is not to scale; the electron orbits are much larger relative to the size of the nucleus.

The only way to account for the particles that reversed direction when they hit the gold foil was to assume that nearly all of the mass, as well as all of the positive charge in each individual gold atom, is concentrated in a tiny center or **nucleus**. When a positively charged alpha particle strikes a nucleus, it reverses direction, much as a cue ball reverses direction when it

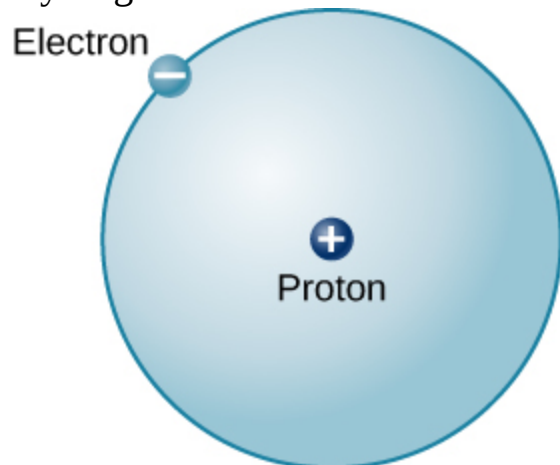
strikes another billiard ball. Rutherford's model placed the other type of charge—the negative electrons—in orbit around this nucleus.

Rutherford's model required that the electrons be in motion. Positive and negative charges attract each other, so stationary electrons would fall into the positive nucleus. Also, because both the electrons and the nucleus are extremely small, most of the atom is empty, which is why nearly all of Rutherford's particles were able to pass right through the gold foil without colliding with anything. Rutherford's model was a very successful explanation of the experiments he conducted, although eventually scientists would discover that even the nucleus itself has structure.

The Atomic Nucleus

The simplest possible atom (and the most common one in the Sun and stars) is hydrogen. The nucleus of ordinary hydrogen contains a single proton. Moving around this proton is a single electron. The mass of an electron is nearly 2000 times smaller than the mass of a proton; the electron carries an amount of charge exactly equal to that of the proton but opposite in sign ([\[link\]](#)). Opposite charges attract each other, so it is an electromagnetic force that holds the proton and electron together, just as gravity is the force that keeps planets in orbit around the Sun.

Hydrogen Atom.

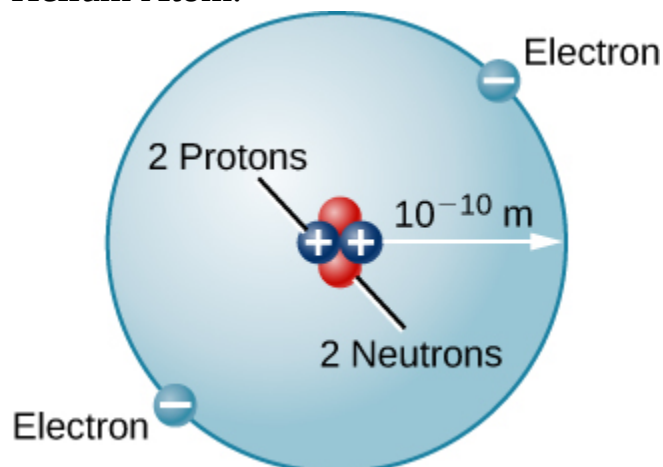


This is a schematic diagram of a hydrogen atom in its lowest energy state, also

called the ground state. The proton and electron have equal but opposite charges, which exert an electromagnetic force that binds the hydrogen atom together. In the illustration, the size of the particles is exaggerated so that you can see them; they are not to scale. They are also shown much closer than they would actually be as it would take more than an entire page to show their actual distance to scale.

There are many other types of atoms in nature. Helium, for example, is the second-most abundant element in the Sun. Helium has two protons in its nucleus instead of the single proton that characterizes hydrogen. In addition, the helium nucleus contains two neutrons, particles with a mass comparable to that of the proton but with no electric charge. Moving around this nucleus are two electrons, so the total net charge of the helium atom is also zero ([link](#)).

Helium Atom.

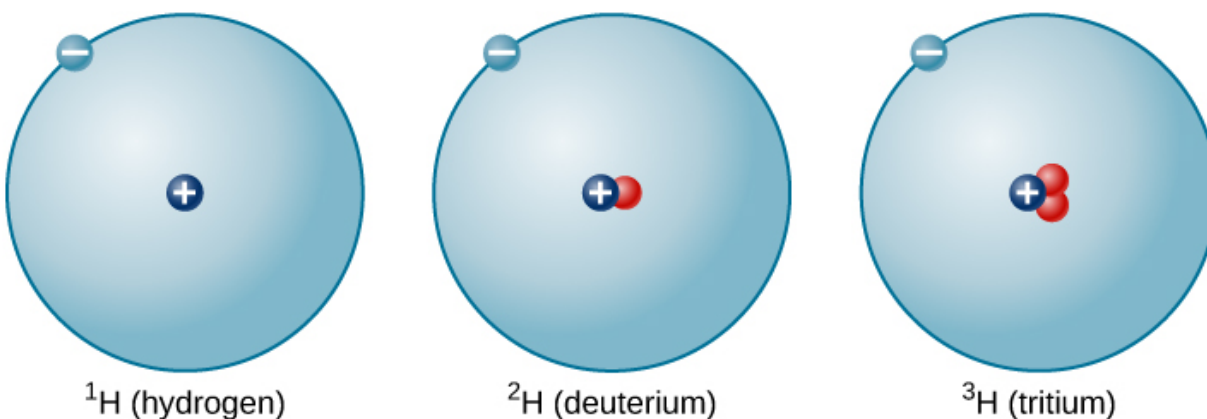


Here we see a schematic diagram of a helium atom in its lowest energy state. Two protons are present in the nucleus of all helium atoms. In the most common variety of helium, the nucleus also contains two neutrons, which have nearly the same mass as the proton but carry no charge. Two electrons orbit the nucleus.

From this description of hydrogen and helium, perhaps you have guessed the pattern for building up all the elements (different types of atoms) that we find in the universe. The type of element is determined by the number of protons in the nucleus of the atom. For example, any atom with six protons is the element carbon, with eight protons is oxygen, with 26 is iron, and with 92 is uranium. On Earth, a typical atom has the same number of electrons as protons, and these electrons follow complex orbital patterns around the nucleus. Deep inside stars, however, it is so hot that the electrons get loose from the nucleus and (as we shall see) lead separate yet productive lives.

The ratio of neutrons to protons increases as the number of protons increases, but each element is unique. The number of neutrons is not necessarily the same for all atoms of a given element. For example, most hydrogen atoms contain no neutrons at all. There are, however, hydrogen atoms that contain one proton and one neutron, and others that contain one proton and two neutrons. The various types of hydrogen nuclei with different numbers of neutrons are called **isotopes** of hydrogen ([\[link\]](#)), and all other elements have isotopes as well. You can think of isotopes as siblings in the same element “family”—closely related but with different characteristics and behaviors.

Isotopes of Hydrogen.



A single proton in the nucleus defines the atom to be hydrogen, but there may be zero, one, or two neutrons. The most common isotope of hydrogen is the one with only a single proton and no neutrons.

Note:

To explore the structure of atoms, go to the [PhET Build an Atom website](#) where you can add protons, neutrons, or electrons to a model and the name of the element you have created will appear. You can also see the net charge, the mass number, whether it is stable or unstable, and whether it is an ion or a neutral atom.

The Bohr Atom

Rutherford's model for atoms has one serious problem. Maxwell's theory of electromagnetic radiation says that when electrons change either speed or the direction of motion, they must emit energy. Orbiting electrons constantly change their direction of motion, so they should emit a constant stream of energy. Applying Maxwell's theory to Rutherford's model, all electrons should spiral into the nucleus of the atom as they lose energy, and this collapse should happen very quickly—in about 10^{-16} seconds.

It was Danish physicist Niels Bohr (1885–1962) who solved the mystery of how electrons remain in orbit. He was trying to develop a model of the atom that would also explain certain regularities observed in the spectrum of hydrogen. He suggested that the spectrum of hydrogen can be understood if we assume that orbits of only certain sizes are possible for the electron. Bohr further assumed that as long as the electron moves in only one of these allowed orbits, it radiates no energy: its energy would change only if it moved from one orbit to another.

This suggestion, in the words of science historian Abraham Pais, was “one of the most audacious hypotheses ever introduced in physics.” If something equivalent were at work in the everyday world, you might find that, as you went for a walk after astronomy class, nature permitted you to walk two steps per minute, five steps per minute, and 12 steps per minute, but no speeds in between. No matter how you tried to move your legs, only certain walking speeds would be permitted. To make things more bizarre, it would take no effort to walk at any one of the allowed speeds, but it would be difficult to change from one speed to another. Luckily, no such rules apply at the level of human behavior. But at the microscopic level of the atom, experiment after experiment has confirmed the validity of Bohr’s strange idea. Bohr’s suggestions became one of the foundations of the new (and much more sophisticated) model of the subatomic world called quantum mechanics.

In Bohr’s model, if the electron moves from one orbit to another closer to the atomic nucleus, it must give up some energy in the form of electromagnetic radiation. If the electron goes from an inner orbit to one farther from the nucleus, however, it requires some additional energy. One way to obtain the necessary energy is to absorb electromagnetic radiation that may be streaming past the atom from an outside source.

A key feature of Bohr’s model is that each of the permitted electron orbits around a given atom has a certain energy value; we therefore can think of each orbit as an **energy level**. To move from one orbit to another (which will have its own specific energy value) requires a change in the electron’s energy—a change determined by the difference between the two energy values. If the electron goes to a lower level, the energy difference will be

given off; if the electron goes to a higher level, the energy difference must be obtained from somewhere else. Each jump (or transition) to a different level has a fixed and definite energy change associated with it.

A crude analogy for this situation might be life in a tower of luxury apartments where the rent is determined by the quality of the view. Such a building has certain, definite numbered levels or floors on which apartments are located. No one can live on floor 5.37 or 22.5. In addition, the rent gets higher as you go up to higher floors. If you want to exchange an apartment on the twentieth floor for one on the second floor, you will not owe as much rent. However, if you want to move from the third floor to the twenty-fifth floor, your rent will increase. In an atom, too, the “cheapest” place for an electron to live is the lowest possible level, and energy is required to move to a higher level.

Here we have one of the situations where it is easier to think of electromagnetic radiation as particles (photons) rather than as waves. As electrons move from one level to another, they give off or absorb little packets of energy. When an electron moves to a higher level, it absorbs a photon of just the right energy (provided one is available). When it moves to a lower level, it emits a photon with the exact amount of energy it no longer needs in its “lower-cost living situation.”

The photon and wave perspectives must be equivalent: light is light, no matter how we look at it. Thus, each photon carries a certain amount of energy that is proportional to the frequency (f) of the wave it represents. The value of its energy (E) is given by the formula

Equation:

$$E = hf$$

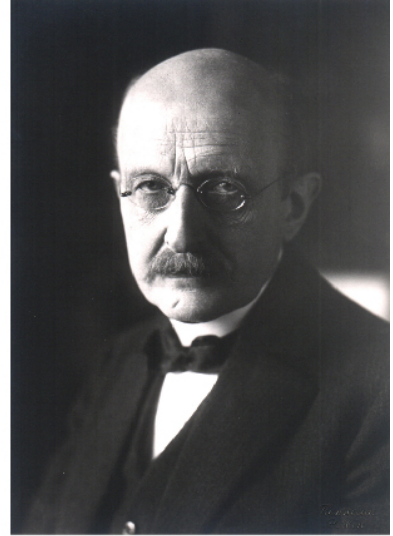
where the constant of proportionality, h , is called Planck’s constant.

The constant is named for Max Planck, the German physicist who was one of the originators of the quantum theory ([link](#)). If metric units are used (that is, if energy is measured in joules and frequency in hertz), then Planck’s constant has the value $h = 6.626 \times 10^{-34}$ joule-seconds (J-s).

Higher-energy photons correspond to higher-frequency waves (which have a shorter wavelength); lower-energy photons are waves of lower frequency. Niels Bohr (1885–1962) and Max Planck (1858–1947).



(a)



(b)

(a) Bohr, shown at his desk in this 1935 photograph, and (b) Planck helped us understand the energy behavior of photons.

To take a specific example, consider a calcium atom inside the Sun's atmosphere in which an electron jumps from a lower level to a higher level. To do this, it needs about 5×10^{-19} joules of energy, which it can conveniently obtain by absorbing a passing photon of that energy coming from deeper inside the Sun. This photon is equivalent to a wave of light whose frequency is about 7.5×10^{14} hertz and whose wavelength is about 3.9×10^{-7} meters (393 nanometers), in the deep violet part of the visible light spectrum. Although it may seem strange at first to switch from picturing light as a photon (or energy packet) to picturing it as a wave, such switching has become second nature to astronomers and can be a handy tool for doing calculations about spectra.

Example:
The Energy of a Photon

Now that we know how to calculate the wavelength and frequency of a photon, we can use this information, along with Planck's constant, to determine how much energy each photon carries. How much energy does a red photon of wavelength 630 nm have?

Solution

First, as we learned earlier, we can find the frequency of the photon:

Equation:

$$f = \frac{c}{\lambda} = \frac{3 \times 10^8 \text{ m/s}}{630 \times 10^{-9} \text{ m}} = 4.8 \times 10^{14} \text{ Hz}$$

Next, we can use Planck's constant to determine the energy (remember that a Hz is the same as 1/s):

Equation:

$$E = hf = 6.626 \times 10^{-34} \text{ J-s} \quad 4.8 \times 10^{14} \text{ Hz}(1/\text{s}) = 3.2 \times 10^{-19} \text{ J}$$

Check Your Learning

What is the energy of a yellow photon with a frequency of $5.5 \times 10^{14} \text{ Hz}$?

Note:

Answer:

$$E = hf = 6.626 \times 10^{-34} \text{ J-s} \quad 5.5 \times 10^{14} \text{ Hz} = 3.6 \times 10^{-19} \text{ J}$$

Key Concepts and Summary

Atoms consist of a nucleus containing one or more positively charged protons. All atoms except hydrogen can also contain one or more neutrons in the nucleus. Negatively charged electrons orbit the nucleus. The number of protons defines an element (hydrogen has one proton, helium has two, and so on) of the atom. Nuclei with the same number of protons but

different numbers of neutrons are different isotopes of the same element. In the Bohr model of the atom, electrons on permitted orbits (or energy levels) don't give off any electromagnetic radiation. But when electrons go from lower levels to higher ones, they must absorb a photon of just the right energy, and when they go from higher levels to lower ones, they give off a photon of just the right energy. The energy of a photon is connected to the frequency of the electromagnetic wave it represents by Planck's formula, $E = hf$.

Glossary

energy level

a particular level, or amount, of energy possessed by an atom or ion above the energy it possesses in its least energetic state; also used to refer to the states of energy an electron can have in an atom

isotope

any of two or more forms of the same element whose atoms have the same number of protons but different numbers of neutrons

nucleus (of an atom)

the massive part of an atom, composed mostly of protons and neutrons, and about which the electrons revolve

Formation of Spectral Lines

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how emission line spectra and absorption line spectra are formed
- Describe what ions are and how they are formed
- Explain how spectral lines and ionization levels in a gas can help us determine its temperature

We can use Bohr's model of the atom to understand how spectral lines are formed. The concept of energy levels for the electron orbits in an atom leads naturally to an explanation of why atoms absorb or emit only specific energies or wavelengths of light.

The Hydrogen Spectrum

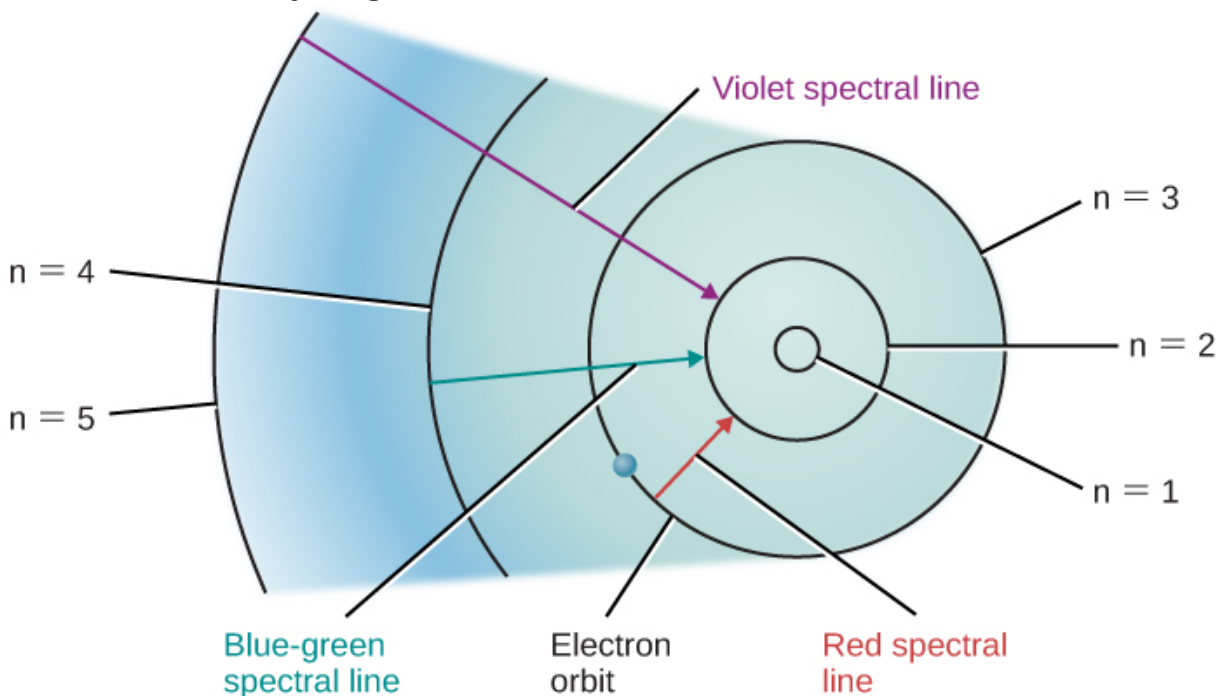
Let's look at the hydrogen atom from the perspective of the Bohr model. Suppose a beam of white light (which consists of photons of all visible wavelengths) shines through a gas of atomic hydrogen. A photon of wavelength 656 nanometers has just the right energy to raise an electron in a hydrogen atom from the second to the third orbit. Thus, as all the photons of different energies (or wavelengths or colors) stream by the hydrogen atoms, photons with *this* particular wavelength can be absorbed by those atoms whose electrons are orbiting on the second level. When they are absorbed, the electrons on the second level will move to the third level, and

a number of the photons of this wavelength and energy will be missing from the general stream of white light.

Other photons will have the right energies to raise electrons from the second to the fourth orbit, or from the first to the fifth orbit, and so on. Only photons with these exact energies can be absorbed. All of the other photons will stream past the atoms untouched. Thus, hydrogen atoms absorb light at only certain wavelengths and produce dark lines at those wavelengths in the spectrum we see.

Suppose we have a container of hydrogen gas through which a whole series of photons is passing, allowing many electrons to move up to higher levels. When we turn off the light source, these electrons “fall” back down from larger to smaller orbits and emit photons of light—but, again, only light of those energies or wavelengths that correspond to the energy difference between permissible orbits. The orbital changes of hydrogen electrons that give rise to some spectral lines are shown in [\[link\]](#).

Bohr Model for Hydrogen.



In this simplified model of a hydrogen atom, the concentric circles shown represent permitted orbits or energy levels. An electron in a hydrogen atom can only exist in one of these energy levels (or states).

The closer the electron is to the nucleus, the more tightly bound the electron is to the nucleus. By absorbing energy, the electron can move to energy levels farther from the nucleus (and even escape if enough energy is absorbed).

Similar pictures can be drawn for atoms other than hydrogen. However, because these other atoms ordinarily have more than one electron each, the orbits of their electrons are much more complicated, and the spectra are more complex as well. For our purposes, the key conclusion is this: *each type of atom has its own unique pattern of electron orbits, and no two sets of orbits are exactly alike*. This means that each type of atom shows its own unique set of spectral lines, produced by electrons moving between its unique set of orbits.

Astronomers and physicists have worked hard to learn the lines that go with each element by studying the way atoms absorb and emit light in laboratories here on Earth. Then they can use this knowledge to identify the elements in celestial bodies. In this way, we now know the chemical makeup of not just any star, but even galaxies of stars so distant that their light started on its way to us long before Earth had even formed.

Energy Levels and Excitation

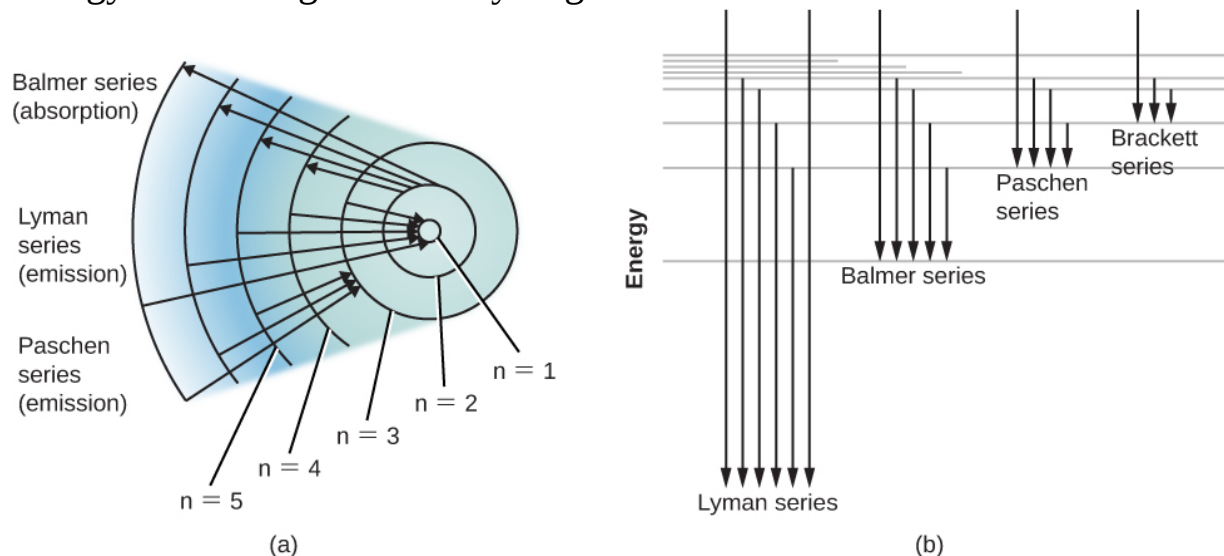
Bohr's model of the hydrogen atom was a great step forward in our understanding of the atom. However, we know today that atoms cannot be represented by quite so simple a picture. For example, the concept of sharply defined electron orbits is not really correct; however, at the level of this introductory course, the notion that only certain discrete energies are allowable for an atom is very useful. The energy levels we have been discussing can be thought of as representing certain average distances of the electron's possible orbits from the atomic nucleus.

Ordinarily, an atom is in the state of lowest possible energy, its **ground state**. In the Bohr model of the hydrogen atom, the ground state corresponds to the electron being in the innermost orbit. An atom can

absorb energy, which raises it to a higher energy level (corresponding, in the simple Bohr picture, to an electron's movement to a larger orbit)—this is referred to as **excitation**. The atom is then said to be in an *excited state*. Generally, an atom remains excited for only a very brief time. After a short interval, typically a hundred-millionth of a second or so, it drops back spontaneously to its ground state, with the simultaneous emission of light. The atom may return to its lowest state in one jump, or it may make the transition in steps of two or more jumps, stopping at intermediate levels on the way down. With each jump, it emits a photon of the wavelength that corresponds to the energy difference between the levels at the beginning and end of that jump.

An energy-level diagram for a hydrogen atom and several possible atomic transitions are shown in [\[link\]](#). When we measure the energies involved as the atom jumps between levels, we find that the transitions to or from the ground state, called the *Lyman series* of lines, result in the emission or absorption of ultraviolet photons. But the transitions to or from the first excited state (labeled $n = 2$ in part (a) of [\[link\]](#)), called the Balmer series, produce emission or absorption in visible light. In fact, it was to explain this Balmer series that Bohr first suggested his model of the atom.

Energy-Level Diagrams for Hydrogen.



(a) Here we follow the emission or absorption of photons by a hydrogen atom according to the Bohr model. Several different series of spectral lines are shown, corresponding to transitions of electrons from

or to certain allowed orbits. Each series of lines that terminates on a specific inner orbit is named for the physicist who studied it. At the top, for example, you see the Balmer series, and arrows show electrons jumping from the second orbit ($n = 2$) to the third, fourth, fifth, and sixth orbits. Each time a “poor” electron from a lower level wants to rise to a higher position in life, it must absorb energy to do so. It can absorb the energy it needs from passing waves (or photons) of light. The next set of arrows (Lyman series) show electrons falling down to the first orbit from different (higher) levels. Each time a “rich” electron goes downward toward the nucleus, it can afford to give off (emit) some energy it no longer needs. (b) At higher and higher energy levels, the levels become more and more crowded together, approaching a limit. The region above the top line represents energies at which the atom is ionized (the electron is no longer attached to the atom). Each series of arrows represents electrons falling from higher levels to lower ones, releasing photons or waves of energy in the process.

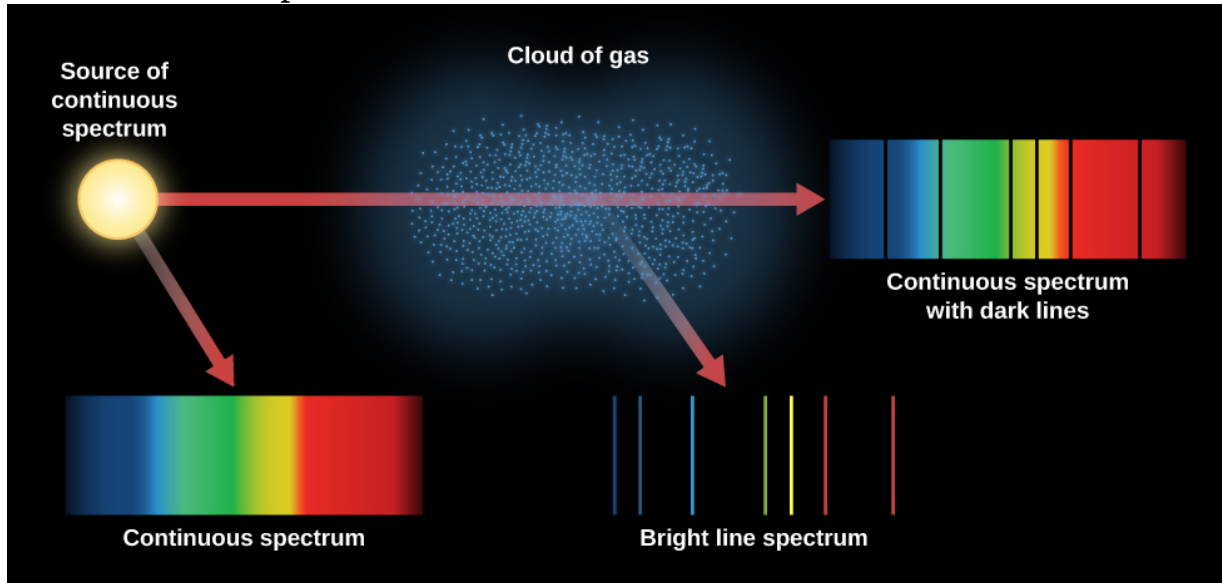
Atoms that have absorbed specific photons from a passing beam of white light and have thus become excited generally de-excite themselves and emit that light again in a very short time. You might wonder, then, why *dark* spectral lines are ever produced. In other words, why doesn’t this reemitted light quickly “fill in” the darker absorption lines?

Imagine a beam of white light coming toward you through some cooler gas. Some of the reemitted light *is* actually returned to the beam of white light you see, but this fills in the absorption lines only to a slight extent. The reason is that the atoms in the gas reemit light *in all directions*, and only a small fraction of the reemitted light is in the direction of the original beam (toward you). In a star, much of the reemitted light actually goes in directions leading back into the star, which does observers outside the star no good whatsoever.

[\[link\]](#) summarizes the different kinds of spectra we have discussed. An incandescent lightbulb produces a continuous spectrum. When that

continuous spectrum is viewed through a thinner cloud of gas, an absorption line spectrum can be seen superimposed on the continuous spectrum. If we look only at a cloud of excited gas atoms (with no continuous source seen behind it), we see that the excited atoms give off an emission line spectrum.

Three Kinds of Spectra.



When we see a lightbulb or other source of continuous radiation, all the colors are present. When the continuous spectrum is seen through a thinner gas cloud, the cloud's atoms produce absorption lines in the continuous spectrum. When the excited cloud is seen without the continuous source behind it, its atoms produce emission lines. We can learn which types of atoms are in the gas cloud from the pattern of absorption or emission lines.

Atoms in a hot gas are moving at high speeds and continually colliding with one another and with any loose electrons. They can be excited (electrons moving to a higher level) and de-excited (electrons moving to a lower level) by these collisions as well as by absorbing and emitting light. The speed of atoms in a gas depends on the temperature. When the temperature is higher, so are the speed and energy of the collisions. The hotter the gas, therefore, the more likely that electrons will occupy the outermost orbits, which

correspond to the highest energy levels. This means that the level where electrons *start* their upward jumps in a gas can serve as an indicator of how hot that gas is. In this way, the absorption lines in a spectrum give astronomers information about the temperature of the regions where the lines originate.

Note:

Use this [simulation](#) to play with a hydrogen atom and see what happens when electrons move to higher levels and then give off photons as they go to a lower level.

Ionization

We have described how certain discrete amounts of energy can be absorbed by an atom, raising it to an excited state and moving one of its electrons farther from its nucleus. If enough energy is absorbed, the electron can be completely removed from the atom—this is called **ionization**. The atom is then said to be ionized. The minimum amount of energy required to remove one electron from an atom in its ground state is called its ionization energy.

Still-greater amounts of energy must be absorbed by the now-ionized atom (called an **ion**) to remove an additional electron deeper in the structure of the atom. Successively greater energies are needed to remove the third, fourth, fifth—and so on—electrons from the atom. If enough energy is available, an atom can become completely ionized, losing all of its electrons. A hydrogen atom, having only one electron to lose, can be ionized only once; a helium atom can be ionized twice; and an oxygen atom up to eight times. When we examine regions of the cosmos where there is a great deal of energetic radiation, such as the neighborhoods where hot young stars have recently formed, we see a lot of ionization going on.

An atom that has become positively ionized has lost a negative charge—the missing electron—and thus is left with a net positive charge. It therefore exerts a strong attraction on any free electron. Eventually, one or more

electrons will be captured and the atom will become neutral (or ionized to one less degree) again. During the electron-capture process, the atom emits one or more photons. Which photons are emitted depends on whether the electron is captured at once to the lowest energy level of the atom or stops at one or more intermediate levels on its way to the lowest available level.

Just as the excitation of an atom can result from a collision with another atom, ion, or electron (collisions with electrons are usually most important), so can ionization. The rate at which such collisional ionizations occur depends on the speeds of the atoms and hence on the temperature of the gas—the hotter the gas, the more of its atoms will be ionized.

The rate at which ions and electrons recombine also depends on their relative speeds—that is, on the temperature. In addition, it depends on the density of the gas: the higher the density, the greater the chance for recapture, because the different kinds of particles are crowded more closely together. From a knowledge of the temperature and density of a gas, it is possible to calculate the fraction of atoms that have been ionized once, twice, and so on. In the Sun, for example, we find that most of the hydrogen and helium atoms in its atmosphere are neutral, whereas most of the calcium atoms, as well as many other heavier atoms, are ionized once.

The energy levels of an ionized atom are entirely different from those of the same atom when it is neutral. Each time an electron is removed from the atom, the energy levels of the ion, and thus the wavelengths of the spectral lines it can produce, change. This helps astronomers differentiate the ions of a given element. Ionized hydrogen, having no electron, can produce no absorption lines.

Key Concepts and Summary

When electrons move from a higher energy level to a lower one, photons are emitted, and an emission line can be seen in the spectrum. Absorption lines are seen when electrons absorb photons and move to higher energy levels. Since each atom has its own characteristic set of energy levels, each is associated with a unique pattern of spectral lines. This allows astronomers to determine what elements are present in the stars and in the

clouds of gas and dust among the stars. An atom in its lowest energy level is in the ground state. If an electron is in an orbit other than the least energetic one possible, the atom is said to be excited. If an atom has lost one or more electrons, it is called an ion and is said to be ionized. The spectra of different ions look different and can tell astronomers about the temperatures of the sources they are observing.

Glossary

excitation

the process of giving an atom or an ion an amount of energy greater than it has in its lowest energy (ground) state

ground state

the lowest energy state of an atom

ion

an atom that has become electrically charged by the addition or loss of one or more electrons

ionization

the process by which an atom gains or loses electrons

The Doppler Effect

LEARNING OBJECTIVES

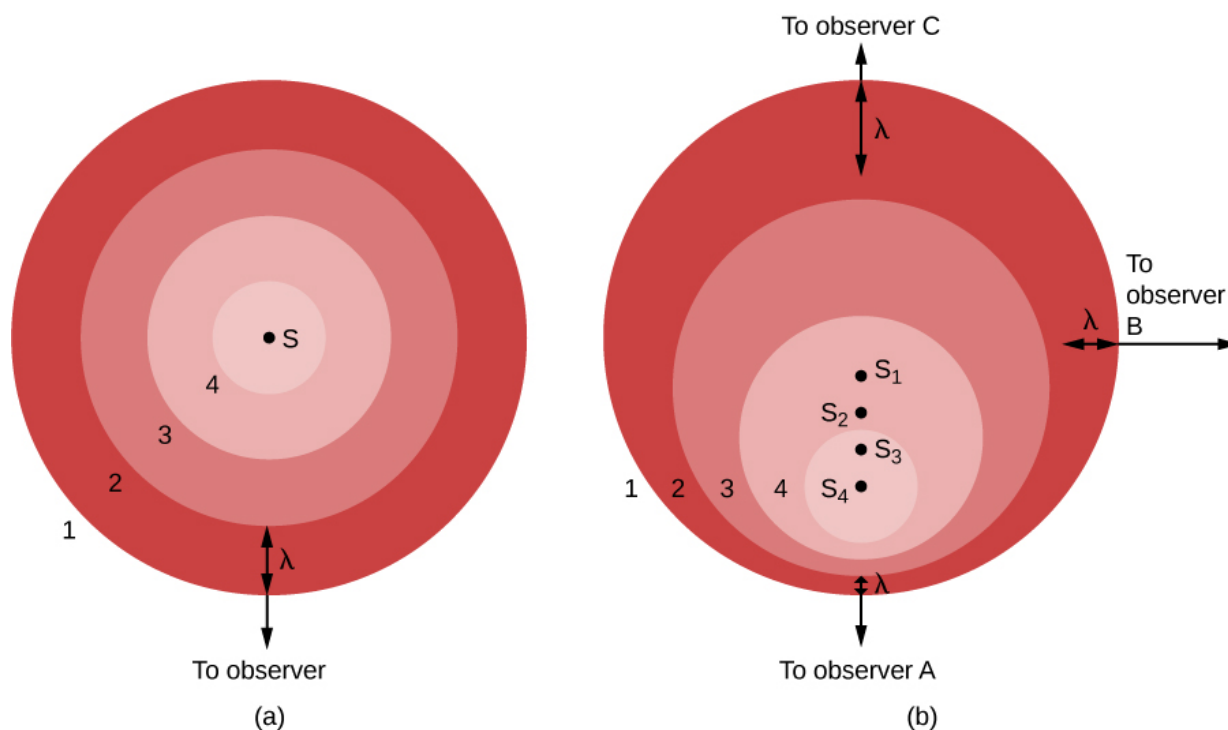
By the end of this section, you will be able to:

- Explain why the spectral lines of photons we observe from an object will change as a result of the object's motion toward or away from us
- Describe how we can use the Doppler effect to deduce how fast astronomical objects are moving through space

The last two sections introduced you to many new concepts, and we hope that through those, you have seen one major idea emerge. Astronomers can learn about the elements in stars and galaxies by decoding the information in their spectral lines. There is a complicating factor in learning how to decode the message of starlight, however. If a star is moving toward or away from us, its lines will be in a slightly different place in the spectrum from where they would be in a star at rest. And most objects in the universe do have some motion relative to the Sun.

Motion Affects Waves

In 1842, Christian Doppler first measured the effect of motion on waves by hiring a group of musicians to play on an open railroad car as it was moving along the track. He then applied what he learned to all waves, including light, and pointed out that if a light source is approaching or receding from the observer, the light waves will be, respectively, crowded more closely together or spread out. The general principle, now known as the **Doppler effect**, is illustrated in [\[link\]](#).
Doppler Effect.



(a) A source, S , makes waves whose numbered crests (1, 2, 3, and 4) wash over a stationary observer. (b) The source S now moves toward observer A and away from observer C . Wave crest 1 was emitted when the source was at position S_1 , crest 2 at position S_2 , and so forth. Observer A sees waves compressed by this motion and sees a blueshift (if the waves are light). Observer C sees the waves stretched out by the motion and sees a redshift. Observer B , whose line of sight is perpendicular to the source's motion, sees no change in the waves (and feels left out).

In part (a) of the figure, the light source (S) is at rest with respect to the observer. The source gives off a series of waves, whose crests we have labeled 1, 2, 3, and 4. The light waves spread out evenly in all directions, like the ripples from a splash in a pond. The crests are separated by a distance, λ , where λ is the wavelength. The observer, who happens to be located in the direction of the bottom of the image, sees the light waves coming nice and evenly, one wavelength apart. Observers located anywhere else would see the same thing.

On the other hand, if the source of light is moving with respect to the observer, as seen in part (b), the situation is more complicated. Between the time one crest is emitted and the next one is ready to come out, the source has moved a bit, toward the bottom of the page. From the point of view of observer A , this motion of the source has decreased the distance between crests—it's squeezing the crests together, this observer might say.

In part (b), we show the situation from the perspective of three observers. The source is seen in four positions, S_1 , S_2 , S_3 , and S_4 , each corresponding to the emission of one wave crest. To observer A , the waves seem to follow one another more closely, at a decreased wavelength and thus increased frequency. (Remember, all light waves travel at the speed of light through empty

space, no matter what. This means that motion cannot affect the speed, but only the wavelength and the frequency. As the wavelength decreases, the frequency must increase. If the waves are shorter, more will be able to move by during each second.)

The situation is not the same for other observers. Let's look at the situation from the point of view of observer *C*, located opposite observer *A* in the figure. For her, the source is moving away from her location. As a result, the waves are not squeezed together but instead are spread out by the motion of the source. The crests arrive with an increased wavelength and decreased frequency. To observer *B*, in a direction at right angles to the motion of the source, no effect is observed. The wavelength and frequency remain the same as they were in part (a) of the figure.

We can see from this illustration that the Doppler effect is produced only by a motion toward or away from the observer, a motion called **radial velocity**. Sideways motion does not produce such an effect. Observers between *A* and *B* would observe some shortening of the light waves for that part of the motion of the source that is along their line of sight. Observers between *B* and *C* would observe lengthening of the light waves that are along their line of sight.

You may have heard the Doppler effect with sound waves. When a train whistle or police siren approaches you and then moves away, you will notice a decrease in the pitch (which is how human senses interpret sound wave frequency) of the sound waves. Compared to the waves at rest, they have changed from slightly more frequent when coming toward you, to slightly less frequent when moving away from you.

Note:

A nice example of this change in the sound of a train whistle can be heard at the end of the classic Beach Boys song "Caroline, No" on their album *Pet Sounds*. To hear this sound, go to this [YouTube](#) version of the song. The sound of the train begins at approximately 2:20.

Color Shifts

When the source of waves moves toward you, the wavelength decreases a bit. If the waves involved are visible light, then the colors of the light change slightly. As wavelength decreases, they shift toward the blue end of the spectrum: astronomers call this a *blueshift* (since the end of the spectrum is really violet, the term should probably be *violetshift*, but blue is a more common color). When the source moves away from you and the wavelength gets longer, we call the change in colors a *redshift*. Because the Doppler effect was first used with visible light in astronomy, the terms "blueshift" and "redshift" became well established. Today, astronomers use these words to describe changes in the wavelengths of radio waves or X-rays as comfortably as they use them to describe changes in visible light.

The greater the motion toward or away from us, the greater the Doppler shift. If the relative motion is entirely along the line of sight, the formula for the Doppler shift of light is

Equation:

$$\frac{\Delta\lambda}{\lambda} = \frac{v}{c}$$

where λ is the wavelength emitted by the source, $\Delta\lambda$ is the difference between λ and the wavelength measured by the observer, c is the speed of light, and v is the relative speed of the observer and the source in the line of sight. The variable v is counted as positive if the velocity is one of recession, and negative if it is one of approach. Solving this equation for the velocity, we find $v = c \times \Delta\lambda/\lambda$.

If a star approaches or recedes from us, the wavelengths of light in its continuous spectrum appear shortened or lengthened, respectively, as do those of the dark lines. However, unless its speed is tens of thousands of kilometers per second, the star does not appear noticeably bluer or redder than normal. The Doppler shift is thus not easily detected in a continuous spectrum and cannot be measured accurately in such a spectrum. The wavelengths of the absorption lines can be measured accurately, however, and their Doppler shift is relatively simple to detect.

Example:

The Doppler Effect

We can use the Doppler effect equation to calculate the radial velocity of an object if we know three things: the speed of light, the original (unshifted) wavelength of the light emitted, and the difference between the wavelength of the emitted light and the wavelength we observe. For particular absorption or emission lines, we usually know exactly what wavelength the line has in our laboratories on Earth, where the source of light is not moving. We can measure the new wavelength with our instruments at the telescope, and so we know the difference in wavelength due to Doppler shifting. Since the speed of light is a universal constant, we can then calculate the radial velocity of the star.

A particular emission line of hydrogen is originally emitted with a wavelength of 656.3 nm from a gas cloud. At our telescope, we observe the wavelength of the emission line to be 656.6 nm. How fast is this gas cloud moving toward or away from Earth?

Solution

Because the light is shifted to a longer wavelength (redshifted), we know this gas cloud is moving away from us. The speed can be calculated using the Doppler shift formula:

Equation:

$$\begin{aligned} v &= c \times \frac{\Delta\lambda}{\lambda} = (3.0 \times 10^8 \text{ m/s}) \left(\frac{0.3 \text{ nm}}{656.3 \text{ nm}} \right) = (3.0 \times 10^8 \text{ m/s}) \left(\frac{0.3 \times 10^{-9} \text{ m}}{656.3 \times 10^{-9} \text{ m}} \right) \\ &= 140,000 \text{ m/s} = 140 \text{ km/s} \end{aligned}$$

Check Your Learning

Suppose a spectral line of hydrogen, normally at 500 nm, is observed in the spectrum of a star to be at 500.1 nm. How fast is the star moving toward or away from Earth?

Note:

Answer:

Because the light is shifted to a longer wavelength, the star is moving away from us:

$$\nu = c \times \frac{\Delta\lambda}{\lambda} = (3.0 \times 10^8 \text{ m/s}) \left(\frac{0.1 \text{ nm}}{500 \text{ nm}} \right) = (3.0 \times 10^8 \text{ m/s}) \left(\frac{0.1 \times 10^{-9} \text{ m}}{500 \times 10^{-9} \text{ m}} \right) = 60,000 \text{ m/s}.$$

Its speed is

60,000 m/s.

You may now be asking: if all the stars are moving and motion changes the wavelength of each spectral line, won't this be a disaster for astronomers trying to figure out what elements are present in the stars? After all, it is the precise wavelength (or color) that tells astronomers which lines belong to which element. And we first measure these wavelengths in containers of gas in our laboratories, which are not moving. If every line in a star's spectrum is now shifted by its motion to a different wavelength (color), how can we be sure which lines and which elements we are looking at in a star whose speed we do not know?

Take heart. This situation sounds worse than it really is. Astronomers rarely judge the presence of an element in an astronomical object by a single line. It is the *pattern* of lines unique to hydrogen or calcium that enables us to determine that those elements are part of the star or galaxy we are observing. The Doppler effect does not change the pattern of lines from a given element—it only shifts the whole pattern slightly toward redder or bluer wavelengths. The shifted pattern is still quite easy to recognize. Best of all, when we do recognize a familiar element's pattern, we get a bonus: the amount the pattern is shifted can enable us to determine the speed of the objects in our line of sight.

The training of astronomers includes much work on learning to decode light (and other electromagnetic radiation). A skillful “decoder” can learn the temperature of a star, what elements are in it, and even its speed in a direction toward us or away from us. That's really an impressive amount of information for stars that are light-years away.

Key Concepts and Summary

If an atom is moving toward us when an electron changes orbits and produces a spectral line, we see that line shifted slightly toward the blue of its normal wavelength in a spectrum. If the atom is moving away, we see the line shifted toward the red. This shift is known as the Doppler effect and can be used to measure the radial velocities of distant objects.

For Further Exploration

Articles

Augensen, H. & Woodbury, J. “The Electromagnetic Spectrum.” *Astronomy* (June 1982): 6.

Darling, D. “Spectral Visions: The Long Wavelengths.” *Astronomy* (August 1984): 16; “The Short Wavelengths.” *Astronomy* (September 1984): 14.

Gingerich, O. “Unlocking the Chemical Secrets of the Cosmos.” *Sky & Telescope* (July 1981): 13.

Stencil, R. et al. “Astronomical Spectroscopy.” *Astronomy* (June 1978): 6.

Websites

Doppler Effect: <http://www.physicsclassroom.com/class/waves/Lesson-3/The-Doppler-Effect>. A shaking bug and the Doppler Effect explained.

Electromagnetic Spectrum: <http://imagine.gsfc.nasa.gov/science/toolbox/emspectrum1.html>. An introduction to the electromagnetic spectrum from NASA’s *Imagine the Universe*; note that you can click the “Advanced” button near the top and get a more detailed discussion.

Rainbows: How They Form and How to See Them: <http://www.livescience.com/30235-rainbows-formation-explainer.html>. By meteorologist and amateur astronomer Joe Rao.

Videos

Doppler Effect: http://www.esa.int/spaceinvideos/Videos/2014/07/Doppler_effect_-_classroom_demonstration_video_VP05. ESA video with Doppler ball demonstration and Doppler effect and satellites (4:48).

How a Prism Works to Make Rainbow Colors: https://www.youtube.com/watch?v=JGqsi_LDUn0. Short video on how a prism bends light to make a rainbow of colors (2:44).

Tour of the Electromagnetic Spectrum: <https://www.youtube.com/watch?v=HPcAWNIVl-8>. NASA Mission Science video tour of the bands of the electromagnetic spectrum (eight short videos).

Introductions to Quantum Mechanics

Ford, Kenneth. *The Quantum World*. 2004. A well-written recent introduction by a physicist/educator.

Gribbin, John. *In Search of Schroedinger’s Cat*. 1984. Clear, very basic introduction to the fundamental ideas of quantum mechanics, by a British physicist and science writer.

Rae, Alastair. *Quantum Physics: A Beginner’s Guide*. 2005. Widely praised introduction by a British physicist.

Collaborative Group Activities

- A. Have your group make a list of all the electromagnetic wave technology you use during a typical day.

- B. How many applications of the Doppler effect can your group think of in everyday life? For example, why would the highway patrol find it useful?
- C. Have members of your group go home and “read” the face of your radio set and then compare notes. If you do not have a radio, research “broadcast radio frequencies” to find answers to the following questions. What do all the words and symbols mean? What frequencies can your radio tune to? What is the frequency of your favorite radio station? What is its wavelength?
- D. If your instructor were to give you a spectrometer, what kind of spectra does your group think you would see from each of the following: (1) a household lightbulb, (2) the Sun, (3) the “neon lights of Broadway,” (4) an ordinary household flashlight, and (5) a streetlight on a busy shopping street?
- E. Suppose astronomers want to send a message to an alien civilization that is living on a planet with an atmosphere very similar to that of Earth’s. This message must travel through space, make it through the other planet’s atmosphere, and be noticeable to the residents of that planet. Have your group discuss what band of the electromagnetic spectrum might be best for this message and why. (Some people, including noted physicist Stephen Hawking, have warned scientists not to send such messages and reveal the presence of our civilization to a possible hostile cosmos. Do you agree with this concern?)

Review Questions

Exercise:

Problem:

What distinguishes one type of electromagnetic radiation from another? What are the main categories (or bands) of the electromagnetic spectrum?

Exercise:

Problem: What is a wave? Use the terms *wavelength* and *frequency* in your definition.

Exercise:

Problem:

Is your textbook the kind of idealized object (described in section on radiation laws) that absorbs all the radiation falling on it? Explain. How about the black sweater worn by one of your classmates?

Exercise:

Problem: Where in an atom would you expect to find electrons? Protons? Neutrons?

Exercise:

Problem:

Explain how emission lines and absorption lines are formed. In what sorts of cosmic objects would you expect to see each?

Exercise:

Problem:

Explain how the Doppler effect works for sound waves and give some familiar examples.

Exercise:

Problem: What kind of motion for a star does not produce a Doppler effect? Explain.

Exercise:

Problem: Describe how Bohr's model used the work of Maxwell.

Exercise:

Problem: Explain why light is referred to as electromagnetic radiation.

Exercise:

Problem:

Explain the difference between radiation as it is used in most everyday language and radiation as it is used in an astronomical context.

Exercise:

Problem: What are the differences between light waves and sound waves?

Exercise:

Problem:

Which type of wave has a longer wavelength: AM radio waves (with frequencies in the kilohertz range) or FM radio waves (with frequencies in the megahertz range)? Explain.

Exercise:

Problem:

Explain why astronomers long ago believed that space must be filled with some kind of substance (the "aether") instead of the vacuum we know it is today.

Exercise:

Problem: Explain what the ionosphere is and how it interacts with some radio waves.

Exercise:

Problem: Which is more dangerous to living things, gamma rays or X-rays? Explain.

Exercise:

Problem:

Explain why we have to observe stars and other astronomical objects from above Earth's atmosphere in order to fully learn about their properties.

Exercise:

Problem:

Explain why hotter objects tend to radiate more energetic photons compared to cooler objects.

Exercise:

Problem: Explain how we can deduce the temperature of a star by determining its color.

Exercise:

Problem:

Explain what dispersion is and how astronomers use this phenomenon to study a star's light.

Exercise:

Problem: Explain why glass prisms disperse light.

Exercise:

Problem: Explain what Joseph Fraunhofer discovered about stellar spectra.

Exercise:

Problem:

Explain how we use spectral absorption and emission lines to determine the composition of a gas.

Exercise:

Problem:

Explain the results of Rutherford's gold foil experiment and how they changed our model of the atom.

Exercise:

Problem:

Is it possible for two different atoms of carbon to have different numbers of neutrons in their nuclei? Explain.

Exercise:

Problem: What are the three isotopes of hydrogen, and how do they differ?

Exercise:

Problem:

Explain how electrons use light energy to move among energy levels within an atom.

Exercise:

Problem:

Explain why astronomers use the term “blueshifted” for objects moving toward us and “redshifted” for objects moving away from us.

Exercise:

Problem:

If spectral line wavelengths are changing for objects based on the radial velocities of those objects, how can we deduce which type of atom is responsible for a particular absorption or emission line?

Thought Questions

Exercise:

Problem:

Make a list of some of the many practical consequences of Maxwell’s theory of electromagnetic waves (television is one example).

Exercise:

Problem: With what type of electromagnetic radiation would you observe:

- A. A star with a temperature of 5800 K?
- B. A gas heated to a temperature of one million K?
- C. A person on a dark night?

Exercise:

Problem:

Why is it dangerous to be exposed to X-rays but not (or at least much less) dangerous to be exposed to radio waves?

Exercise:

Problem:

Go outside on a clear night, wait 15 minutes for your eyes to adjust to the dark, and look carefully at the brightest stars. Some should look slightly red and others slightly blue. The primary factor that determines the color of a star is its temperature. Which is hotter: a blue star or a red one? Explain

Exercise:

Problem:

Water faucets are often labeled with a red dot for hot water and a blue dot for cold. Given Wien's law, does this labeling make sense?

Exercise:

Problem:

Suppose you are standing at the exact center of a park surrounded by a circular road. An ambulance drives completely around this road, with siren blaring. How does the pitch of the siren change as it circles around you?

Exercise:

Problem:

How could you measure Earth's orbital speed by photographing the spectrum of a star at various times throughout the year? (Hint: Suppose the star lies in the plane of Earth's orbit.)

Exercise:

Problem:

Astronomers want to make maps of the sky showing sources of X-rays or gamma rays. Explain why those X-rays and gamma rays must be observed from above Earth's atmosphere.

Exercise:

Problem:

The greenhouse effect can be explained easily if you understand the laws of blackbody radiation. A greenhouse gas blocks the transmission of infrared light. Given that the incoming light to Earth is sunlight with a characteristic temperature of 5800 K (which peaks in the visible part of the spectrum) and the outgoing light from Earth has a characteristic temperature of about 300 K (which peaks in the infrared part of the spectrum), explain how greenhouse gases cause Earth to warm up. As part of your answer, discuss that greenhouse gases block both incoming and outgoing infrared light. Explain why these two effects don't simply cancel each other, leading to no net temperature change.

Exercise:

Problem:

An idealized radiating object does not reflect or scatter any radiation but instead absorbs all of the electromagnetic energy that falls on it. Can you explain why astronomers call such an object a blackbody? Keep in mind that even stars, which shine brightly in a variety of colors, are considered blackbodies. Explain why.

Exercise:

Problem:

Why are ionized gases typically only found in very high-temperature environments?

Exercise:

Problem: Explain why each element has a unique spectrum of absorption or emission lines.

Figuring for Yourself

Exercise:

Problem:

What is the wavelength of the carrier wave of a campus radio station, broadcasting at a frequency of 97.2 MHz (million cycles per second or million hertz)?

Exercise:

Problem:

What is the frequency of a red laser beam, with a wavelength of 670 nm, which your astronomy instructor might use to point to slides during a lecture on galaxies?

Exercise:

Problem:

You go to a dance club to forget how hard your astronomy midterm was. What is the frequency of a wave of ultraviolet light coming from a blacklight in the club, if its wavelength is 150 nm?

Exercise:

Problem: What is the energy of the photon with the frequency you calculated in [\[link\]](#)?

Exercise:

Problem:

If the emitted infrared radiation from Pluto, has a wavelength of maximum intensity at 75,000 nm, what is the temperature of Pluto assuming it follows Wien's law?

Exercise:

Problem:

What is the temperature of a star whose maximum light is emitted at a wavelength of 290 nm?

Glossary

Doppler effect

the apparent change in wavelength or frequency of the radiation from a source due to its relative motion away from or toward the observer

radial velocity

motion toward or away from the observer; the component of relative velocity that lies in the line of sight

Earth's Crust

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Denote the primary types of rock that constitute Earth's crust
- Explain the theory of plate tectonics
- Describe the difference between rift and subduction zones
- Describe the relationship between fault zones and mountain building
- Explain the various types of volcanic activity occurring on Earth

Let us now examine our planet's outer layers in more detail. Earth's crust is a dynamic place. Volcanic eruptions, erosion, and large-scale movements of the continents rework the surface of our planet constantly. Geologically, ours is the most active planet. Many of the geological processes described in this section have taken place on other planets as well, but usually in their distant pasts. Some of the moons of the giant planets also have impressive activity levels. For example, Jupiter's moon Io has a remarkable number of active volcanoes.

Composition of the Crust

Earth's crust is largely made up of oceanic basalt and continental granite. These are both **igneous rock**, the term used for any rock that has cooled from a molten state. All volcanically produced rock is igneous ([\[link\]](#)).
Formation of Igneous Rock as Liquid Lava Cools and Freezes.



This is a lava flow from a basaltic eruption. Basaltic lava flows quickly and can move easily over distances of more than 20 kilometers. (credit: USGS)

Two other kinds of rock are familiar to us on Earth, although it turns out that neither is common on other planets. **Sedimentary rocks** are made of fragments of igneous rock or the shells of living organisms deposited by wind or water and cemented together without melting. On Earth, these rocks include the common sandstones, shales, and limestones.

Metamorphic rocks are produced when high temperature or pressure alters igneous or sedimentary rock physically or chemically (the word *metamorphic* means “changed in form”). Metamorphic rocks are produced on Earth because geological activity carries surface rocks down to considerable depths and then brings them back up to the surface. Without such activity, these changed rocks would not exist at the surface.

There is a fourth very important category of rock that can tell us much about the early history of the planetary system: **primitive rock**, which has largely escaped chemical modification by heating. Primitive rock represents the original material out of which the planetary system was made. No primitive material is left on Earth because the entire planet was heated early

in its history. To find primitive rock, we must look to smaller objects such as comets, asteroids, and small planetary moons. We can sometimes see primitive rock in samples that fall to Earth from these smaller objects.

A block of quartzite on Earth is composed of materials that have gone through all four of these states. Beginning as primitive material before Earth was born, it was heated in the early Earth to form igneous rock, transformed chemically and redeposited (perhaps many times) to form sedimentary rock, and finally changed several kilometers below Earth's surface into the hard, white metamorphic stone we see today.

Plate Tectonics

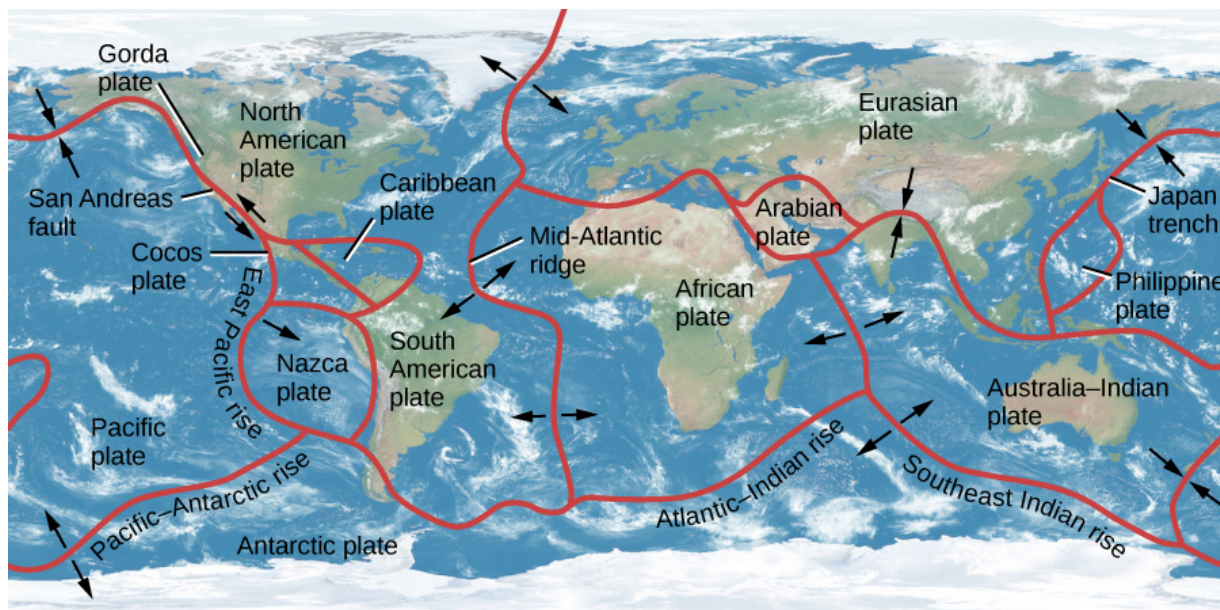
Geology is the study of Earth's crust and the processes that have shaped its surface throughout history. (Although *geo-* means “related to Earth,” astronomers and planetary scientists also talk about the geology of other planets.) Heat escaping from the interior provides energy for the formation of our planet's mountains, valleys, volcanoes, and even the continents and ocean basins themselves. But not until the middle of the twentieth century did geologists succeed in understanding just how these landforms are created.

Plate tectonics is a theory that explains how slow motions within the mantle of Earth move large segments of the crust, resulting in a gradual “drifting” of the continents as well as the formation of mountains and other large-scale geological features. Plate tectonics is a concept as basic to geology as evolution by natural selection is to biology or gravity is to understanding the orbits of planets. Looking at it from a different perspective, plate tectonics is a mechanism for Earth to transport heat efficiently from the interior, where it has accumulated, out to space. It is a cooling system for the planet. All planets develop a heat transfer process as they evolve; mechanisms may differ from that on Earth as a result of chemical makeup and other constraints.

Earth's crust and upper mantle (to a depth of about 60 kilometers) are divided into about a dozen tectonic plates that fit together like the pieces of a jigsaw puzzle ([link](#)). In some places, such as the Atlantic Ocean, the

plates are moving apart; in others, such as off the western coast of South America, they are being forced together. The power to move the plates is provided by slow **convection** of the mantle, a process by which heat escapes from the interior through the upward flow of warmer material and the slow sinking of cooler material. (Convection, in which energy is transported from a warm region, such as the interior of Earth, to a cooler region, such as the upper mantle, is a process we encounter often in astronomy—in stars as well as planets. It is also important in boiling water for coffee while studying for astronomy exams.)

Earth's Continental Plates.



This map shows the major plates into which the crust of Earth is divided. Arrows indicate the motion of the plates at average speeds of 4 to 5 centimeters per year, similar to the rate at which your hair grows.

Note:

The US Geological Survey provides a [map of recent earthquakes](#) and shows the boundaries of the tectonic plates and where earthquakes occur in

relation to these boundaries. You can look close-up at the United States or zoom out for a global view.

As the plates slowly move, they bump into each other and cause dramatic changes in Earth's crust over time. Four basic kinds of interactions between crustal plates are possible at their boundaries: (1) they can pull apart, (2) one plate can burrow under another, (3) they can slide alongside each other, or (4) they can jam together. Each of these activities is important in determining the geology of Earth.

Note:

Alfred Wegener: Catching the Drift of Plate Tectonics

When studying maps or globes of Earth, many students notice that the coast of North and South America, with only minor adjustments, could fit pretty well against the coast of Europe and Africa. It seems as if these great landmasses could once have been together and then were somehow torn apart. The same idea had occurred to others (including Francis Bacon as early as 1620), but not until the twentieth century could such a proposal be more than speculation. The scientist who made the case for continental drift in 1920 was a German meteorologist and astronomer named Alfred Wegener ([\[link\]](#)).

Alfred Wegener (1880–1930).



Wegener proposed a scientific theory for the slow shifting of the continents.

Born in Berlin in 1880, Wegener was, from an early age, fascinated by Greenland, the world's largest island, which he dreamed of exploring. He studied at the universities in Heidelberg, Innsbruck, and Berlin, receiving a doctorate in astronomy by reexamining thirteenth-century astronomical tables. But, his interests turned more and more toward Earth, particularly its weather. He carried out experiments using kites and balloons, becoming so accomplished that he and his brother set a world record in 1906 by flying for 52 hours in a balloon.

Wegener first conceived of continental drift in 1910 while examining a world map in an atlas, but it took 2 years for him to assemble sufficient data to propose the idea in public. He published the results in book form in 1915. Wegener's evidence went far beyond the congruence in the shapes of the continents. He proposed that the similarities between fossils found only in South America and Africa indicated that these two continents were joined at one time. He also showed that resemblances among living animal species on different continents could best be explained by assuming that the continents were once connected in a supercontinent he called *Pangaea* (from Greek elements meaning "all land").

Wegener's suggestion was met with a hostile reaction from most scientists. Although he had marshaled an impressive list of arguments for his hypothesis, he was missing a *mechanism*. No one could explain *how* solid continents could drift over thousands of miles. A few scientists were sufficiently impressed by Wegener's work to continue searching for additional evidence, but many found the notion of moving continents too revolutionary to take seriously. Developing an understanding of the mechanism (plate tectonics) would take decades of further progress in geology, oceanography, and geophysics.

Wegener was disappointed in the reception of his suggestion, but he continued his research and, in 1924, he was appointed to a special meteorology and geophysics professorship created especially for him at the University of Graz (where he was, however, ostracized by most of the geology faculty). Four years later, on his fourth expedition to his beloved Greenland, he celebrated his fiftieth birthday with colleagues and then set off on foot toward a different camp on the island. He never made it; he was found a few days later, dead of an apparent heart attack.

Critics of science often point to the resistance to the continental drift hypothesis as an example of the flawed way that scientists regard new ideas. (Many people who have advanced crackpot theories have claimed that they are being ridiculed unjustly, just as Wegener was.) But we think there is a more positive light in which to view the story of Wegener's suggestion. Scientists in his day maintained a skeptical attitude because they needed more evidence and a clear mechanism that would fit what they understood about nature. Once the evidence and the mechanism were clear, Wegener's hypothesis quickly became the centerpiece of our view of a dynamic Earth.

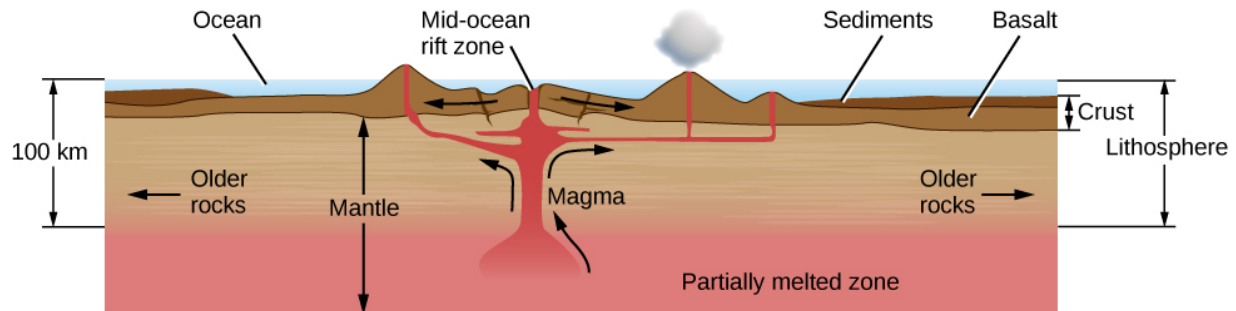
Note:

See how the [drift of the continents](#) has changed the appearance of our planet's crust.

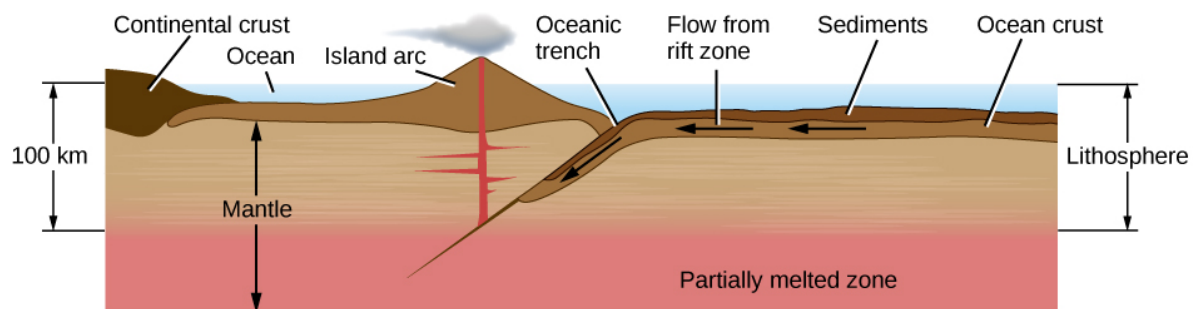
Rift and Subduction Zones

Plates pull apart from each other along **rift zones**, such as the Mid-Atlantic ridge, driven by upwelling currents in the mantle ([link](#)). A few rift zones are found on land. The best known is the central African rift—an area where the African continent is slowly breaking apart. Most rift zones, however, are in the oceans. Molten rock rises from below to fill the space between the receding plates; this rock is basaltic lava, the kind of igneous rock that forms most of the ocean basins.

Rift Zone and Subduction Zone.



Rift zone



Subduction zone

Rift and subduction zones are the regions (mostly beneath the oceans) where new crust is formed and old crust is destroyed as part of the cycle of plate tectonics.

From a knowledge of how the seafloor is spreading, we can calculate the average age of the oceanic crust. About 60,000 kilometers of active rifts have been identified, with average separation rates of about 4 centimeters per year. The new area added to Earth each year is about 2 square kilometers, enough to renew the entire oceanic crust in a little more than 100 million years. This is a very short interval in geological time—less than 3% of the age of Earth. The present ocean basins thus turn out to be among the youngest features on our planet.

As new crust is added to Earth, the old crust must go somewhere. When two plates come together, one plate is often forced beneath another in what is called a **subduction** zone ([\[link\]](#)). In general, the thick continental masses cannot be subducted, but the thinner oceanic plates can be rather readily thrust down into the upper mantle. Often a subduction zone is marked by an ocean trench; a fine example of this type of feature is the deep Japan trench along the coast of Asia. The subducted plate is forced down into regions of high pressure and temperature, eventually melting several hundred kilometers below the surface. Its material is recycled into a downward-flowing convection current, ultimately balancing the flow of material that rises along rift zones. The amount of crust destroyed at subduction zones is approximately equal to the amount formed at rift zones.

All along the subduction zone, earthquakes and volcanoes mark the death throes of the plate. Some of the most destructive earthquakes in history have taken place along subduction zones, including the 1923 Yokohama earthquake and fire that killed 100,000 people, the 2004 Sumatra earthquake and tsunami that killed more than 200,000 people, and the 2011 Tohoku earthquake that resulted in the meltdown of three nuclear power reactors in Japan.

Fault Zones and Mountain Building

Along much of their length, the crustal plates slide parallel to each other. These plate boundaries are marked by cracks or **faults**. Along active fault

zones, the motion of one plate with respect to the other is several centimeters per year, about the same as the spreading rates along rifts.

One of the most famous faults is the San Andreas Fault in California, which lies at the boundary between the Pacific plate and the North American plate ([link](#)). This fault runs from the Gulf of California to the Pacific Ocean northwest of San Francisco. The Pacific plate, to the west, is moving northward, carrying Los Angeles, San Diego, and parts of the southern California coast with it. In several million years, Los Angeles may be an island off the coast of San Francisco.

San Andreas Fault.



We see part of a very active region in California where one crustal plate is sliding sideways with respect to the other. The fault is marked by the valley running up the right side of the photo. Major slippages along this fault can produce extremely destructive earthquakes. (credit: John Wiley)

Unfortunately for us, the motion along fault zones does not take place smoothly. The creeping motion of the plates against each other builds up stresses in the crust that are released in sudden, violent slippages that generate earthquakes. Because the average motion of the plates is constant, the longer the interval between earthquakes, the greater the stress and the more energy released when the surface finally moves.

For example, the part of the San Andreas Fault near the central California town of Parkfield has slipped every 25 years or so during the past century, moving an average of about 1 meter each time. In contrast, the average interval between major earthquakes in the Los Angeles region is about 150 years, and the average motion is about 7 meters. The last time the San Andreas fault slipped in this area was in 1857; tension has been building ever since, and sometime soon it is bound to be released. Sensitive instruments placed within the Los Angeles basin show that the basin is distorting and contracting in size as these tremendous pressures build up beneath the surface.

Example:**Fault Zones and Plate Motion**

After scientists mapped the boundaries between tectonic plates in Earth's crust and measured the annual rate at which the plates move (which is about 5 cm/year), we could estimate quite a lot about the rate at which the geology of Earth is changing. As an example, let's suppose that the next slippage along the San Andreas Fault in southern California takes place in the year 2017 and that it completely relieves the accumulated strain in this region. How much slippage is required for this to occur?

Solution

The speed of motion of the Pacific plate relative to the North American plate is 5 cm/y. That's 500 cm (or 5 m) per century. The last southern California earthquake was in 1857. The time from 1857 to 2017 is 160 y, or 1.6 centuries, so the slippage to relieve the strain completely

would be

$5 \text{ m/century} \times 1.6 \text{ centuries} = 8.0 \text{ m.}$

Check Your Learning

If the next major southern California earthquake occurs in 2047 and only relieves one-half of the accumulated strain, how much slippage will occur?

Note:

Answer:

The difference in time from 1857 to 2047 is 190 y, or 1.9 centuries.

Because only half the strain is released, this is equivalent to half the annual rate of motion. The total slippage comes to

$0.5 \times 5 \text{ m/century} \times 1.9 \text{ centuries} = 4.75 \text{ m.}$

When two continental masses are moving on a collision course, they push against each other under great pressure. Earth buckles and folds, dragging some rock deep below the surface and raising other folds to heights of many kilometers. This is the way many, but not all, of the mountain ranges on Earth were formed. The Alps, for example, are a result of the African plate bumping into the Eurasian plate. As we will see, however, quite different processes produced the mountains on other planets.

Once a mountain range is formed by upthrusting of the crust, its rocks are subject to erosion by water and ice. The sharp peaks and serrated edges have little to do with the forces that make the mountains initially. Instead, they result from the processes that tear down mountains. Ice is an especially effective sculptor of rock ([\[link\]](#)). In a world without moving ice or running water (such as the Moon or Mercury), mountains remain smooth and dull. **Mountains on Earth.**



The Torres del Paine are a young region of Earth's crust where sharp mountain peaks are being sculpted by glaciers. We owe the beauty of our young, steep mountains to the erosion by ice and water. (credit: David Morrison)

Volcanoes

Volcanoes mark locations where lava rises to the surface. One example is mid ocean ridges, which are long undersea mountain ranges formed by lava rising from Earth's mantle at plate boundaries. A second major kind of volcanic activity is associated with subduction zones, and volcanoes sometimes also appear in regions where continental plates are colliding. In each case, the volcanic activity gives us a way to sample some of the material from deeper within our planet.

Other volcanic activity occurs above mantle "hot spots"—areas far from plate boundaries where heat is nevertheless rising from the interior of Earth.

One of the best-known hot spot is under the island of Hawaii, where it currently supplies the heat to maintain three active volcanoes, two on land and one under the ocean. The Hawaii hot spot has been active for at least 100 million years. As Earth's plates have moved during that time, the hot spot has generated a 3500-kilometer-long chain of volcanic islands. The tallest Hawaiian volcanoes are among the largest individual mountains on Earth, more than 100 kilometers in diameter and rising 9 kilometers above the ocean floor. One of the Hawaiian volcanic mountains, the now-dormant Mauna Kea, has become one of the world's great sites for doing astronomy.

Note:

The US Geological Service provides [an interactive map](#) of the famous “ring of fire,” which is the chain of volcanoes surrounding the Pacific Ocean, and shows the Hawaiian “hot spot” enclosed within.

Not all volcanic eruptions produce mountains. If lava flows rapidly from long cracks, it can spread out to form lava plains. The largest known terrestrial eruptions, such as those that produced the Snake River basalts in the northwestern United States or the Deccan plains in India, are of this type. Similar lava plains are found on the Moon and the other terrestrial planets.

Key Concepts and Summary

Terrestrial rocks can be classified as igneous, sedimentary, or metamorphic. A fourth type, primitive rock, is not found on Earth. Our planet's geology is dominated by plate tectonics, in which crustal plates move slowly in response to mantle convection. The surface expression of plate tectonics includes continental drift, recycling of the ocean floor, mountain building, rift zones, subduction zones, faults, earthquakes, and volcanic eruptions of lava from the interior.

Glossary

convection

movement caused within a gas or liquid by the tendency of hotter, and therefore less dense material, to rise and colder, denser material to sink under the influence of gravity, which consequently results in transfer of heat

fault

in geology, a crack or break in the crust of a planet along which slippage or movement can take place, accompanied by seismic activity

igneous rock

rock produced by cooling from a molten state

metamorphic rock

rock produced by physical and chemical alteration (without melting) under high temperature and pressure

plate tectonics

the motion of segments or plates of the outer layer of a planet over the underlying mantle

primitive rock

rock that has not experienced great heat or pressure and therefore remains representative of the original condensed materials from the solar nebula

rift zone

in geology, a place where the crust is being torn apart by internal forces generally associated with the injection of new material from the mantle and with the slow separation of tectonic plates

sedimentary rock

rock formed by the deposition and cementing of fine grains of material, such as pieces of igneous rock or the shells of living things

subduction

the sideways and downward movement of the edge of a plate of Earth's crust into the mantle beneath another plate

volcano

a place where material from a planet's mantle erupts on its surface

Earth's Atmosphere

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Differentiate between Earth's various atmospheric layers
- Describe the chemical composition and possible origins of our atmosphere
- Explain the difference between weather and climate

We live at the bottom of the ocean of air that envelops our planet. The atmosphere, weighing down upon Earth's surface under the force of gravity, exerts a pressure at sea level that scientists define as 1 **bar** (a term that comes from the same root as *barometer*, an instrument used to measure atmospheric pressure). A bar of pressure means that each square centimeter of Earth's surface has a weight equivalent to 1.03 kilograms pressing down on it. Humans have evolved to live at this pressure; make the pressure a lot lower or higher and we do not function well.

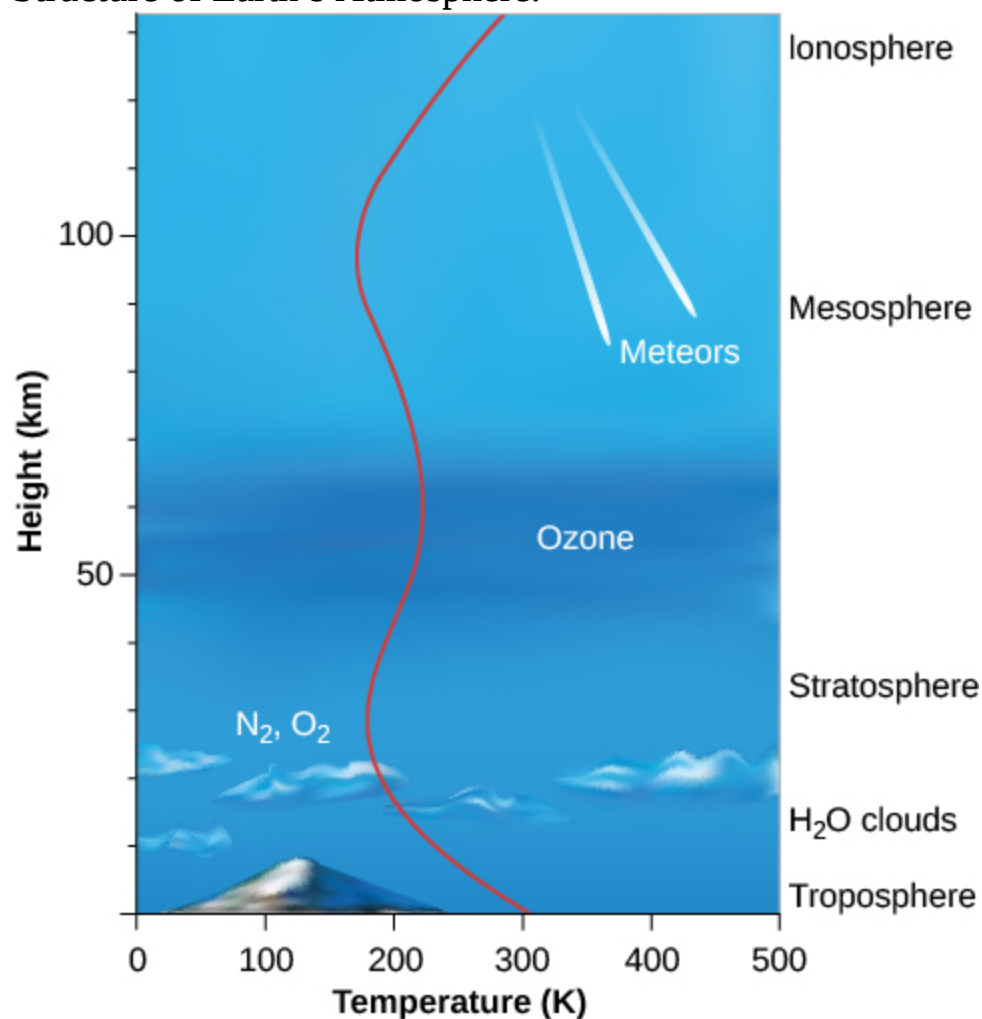
The total mass of Earth's atmosphere is about 5×10^{18} kilograms. This sounds like a large number, but it is only about a millionth of the total mass of Earth. The atmosphere represents a smaller fraction of Earth than the fraction of your mass represented by the hair on your head.

Structure of the Atmosphere

The structure of the atmosphere is illustrated in [\[link\]](#). Most of the atmosphere is concentrated near the surface of Earth, within about the

bottom 10 kilometers where clouds form and airplanes fly. Within this region—called the **troposphere**—warm air, heated by the surface, rises and is replaced by descending currents of cooler air; this is an example of convection. This circulation generates clouds and wind. Within the troposphere, temperature decreases rapidly with increasing elevation to values near 50 °C below freezing at its upper boundary, where the **stratosphere** begins. Most of the stratosphere, which extends to about 50 kilometers above the surface, is cold and free of clouds.

Structure of Earth's Atmosphere.



Height increases up the left side of the diagram, and the names of the different atmospheric layers are shown at the right. In the upper ionosphere, ultraviolet radiation from the Sun can strip electrons from their atoms, leaving the atmosphere ionized. The curving

red line shows the temperature (see the scale on the x -axis).

Near the top of the stratosphere is a layer of **ozone** (O_3), a heavy form of oxygen with three atoms per molecule instead of the usual two. Because ozone is a good absorber of ultraviolet light, it protects the surface from some of the Sun's dangerous ultraviolet radiation, making it possible for life to exist on Earth. The breakup of ozone adds heat to the stratosphere, reversing the decreasing temperature trend in the troposphere. Because ozone is essential to our survival, we reacted with justifiable concern to evidence that became clear in the 1980s that atmospheric ozone was being destroyed by human activities. By international agreement, the production of industrial chemicals that cause ozone depletion, called chlorofluorocarbons, or CFCs, has been phased out. As a result, ozone loss has stopped and the "ozone hole" over the Antarctic is shrinking gradually. This is an example of how concerted international action can help maintain the habitability of Earth.

Note:

Visit NASA's scientific visualization studio for a [short video](#) of what would have happened to Earth's ozone layer by 2065 if CFCs had not been regulated.

At heights above 100 kilometers, the atmosphere is so thin that orbiting satellites can pass through it with very little friction. Many of the atoms are ionized by the loss of an electron, and this region is often called the ionosphere. At these elevations, individual atoms can occasionally escape completely from the gravitational field of Earth. There is a continuous, slow leaking of atmosphere—especially of lightweight atoms, which move faster than heavy ones. Earth's atmosphere cannot, for example, hold on for long to hydrogen or helium, which escape into space. Earth is not the only planet to experience atmosphere leakage. Atmospheric leakage also created Mars'

thin atmosphere. Venus' dry atmosphere evolved because its proximity to the Sun vaporized and dissociated any water, with the component gases lost to space.

Atmospheric Composition and Origin

At Earth's surface, the atmosphere consists of 78% nitrogen (N_2), 21% oxygen (O_2), and 1% argon (Ar), with traces of water vapor (H_2O), carbon dioxide (CO_2), and other gases. Variable amounts of dust particles and water droplets are also found suspended in the air.

A complete census of Earth's volatile materials, however, should look at more than the gas that is now present. *Volatile* materials are those that evaporate at a relatively low temperature. If Earth were just a little bit warmer, some materials that are now liquid or solid might become part of the atmosphere. Suppose, for example, that our planet were heated to above the boiling point of water (100°C , or 373 K); that's a large change for humans, but a small change compared to the range of possible temperatures in the universe. At 100°C , the oceans would boil and the resulting water vapor would become a part of the atmosphere.

To estimate how much water vapor would be released, note that there is enough water to cover the entire Earth to a depth of about 300 meters. Because the pressure exerted by 10 meters of water is equal to about 1 bar, the average pressure at the ocean floor is about 300 bars. Water weighs the same whether in liquid or vapor form, so if the oceans boiled away, the atmospheric pressure of the water would still be 300 bars. Water would therefore greatly dominate Earth's atmosphere, with nitrogen and oxygen reduced to the status of trace constituents.

On a warmer Earth, another source of additional atmosphere would be found in the sedimentary carbonate rocks of the crust. These minerals contain abundant carbon dioxide. If all these rocks were heated, they would release about 70 bars of CO_2 , far more than the current CO_2 pressure of only 0.0005 bar. Thus, the atmosphere of a warm Earth would be dominated by water vapor and carbon dioxide, with a surface pressure nearing 400 bars.

Several lines of evidence show that the composition of Earth's atmosphere has changed over our planet's history. Scientists can infer the amount of atmospheric oxygen, for example, by studying the chemistry of minerals that formed at various times. We examine this issue in more detail later in this chapter.

Today we see that CO₂, H₂O, sulfur dioxide (SO₂), and other gases are released from deeper within Earth through the action of volcanoes. (For CO₂, the primary source today is the burning of fossil fuels, which releases far more CO₂ than that from volcanic eruptions.) Much of this apparently new gas, however, is recycled material that has been subducted through plate tectonics. But where did our planet's original atmosphere come from?

Three possibilities exist for the original source of Earth's atmosphere and oceans: (1) the atmosphere could have been formed with the rest of Earth as it accumulated from debris left over from the formation of the Sun; (2) it could have been released from the interior through volcanic activity, subsequent to the formation of Earth; or (3) it may have been derived from impacts by comets and asteroids from the outer parts of the solar system. Current evidence favors a combination of the interior and impact sources.

Weather and Climate

All planets with atmospheres have *weather*, which is the name we give to the circulation of the atmosphere. The energy that powers the weather is derived primarily from the sunlight that heats the surface. Both the rotation of the planet and slower seasonal changes cause variations in the amount of sunlight striking different parts of Earth. The atmosphere and oceans redistribute the heat from warmer to cooler areas. Weather on any planet represents the response of its atmosphere to changing inputs of energy from the Sun (see [\[link\]](#) for a dramatic example).

Storm from Space.



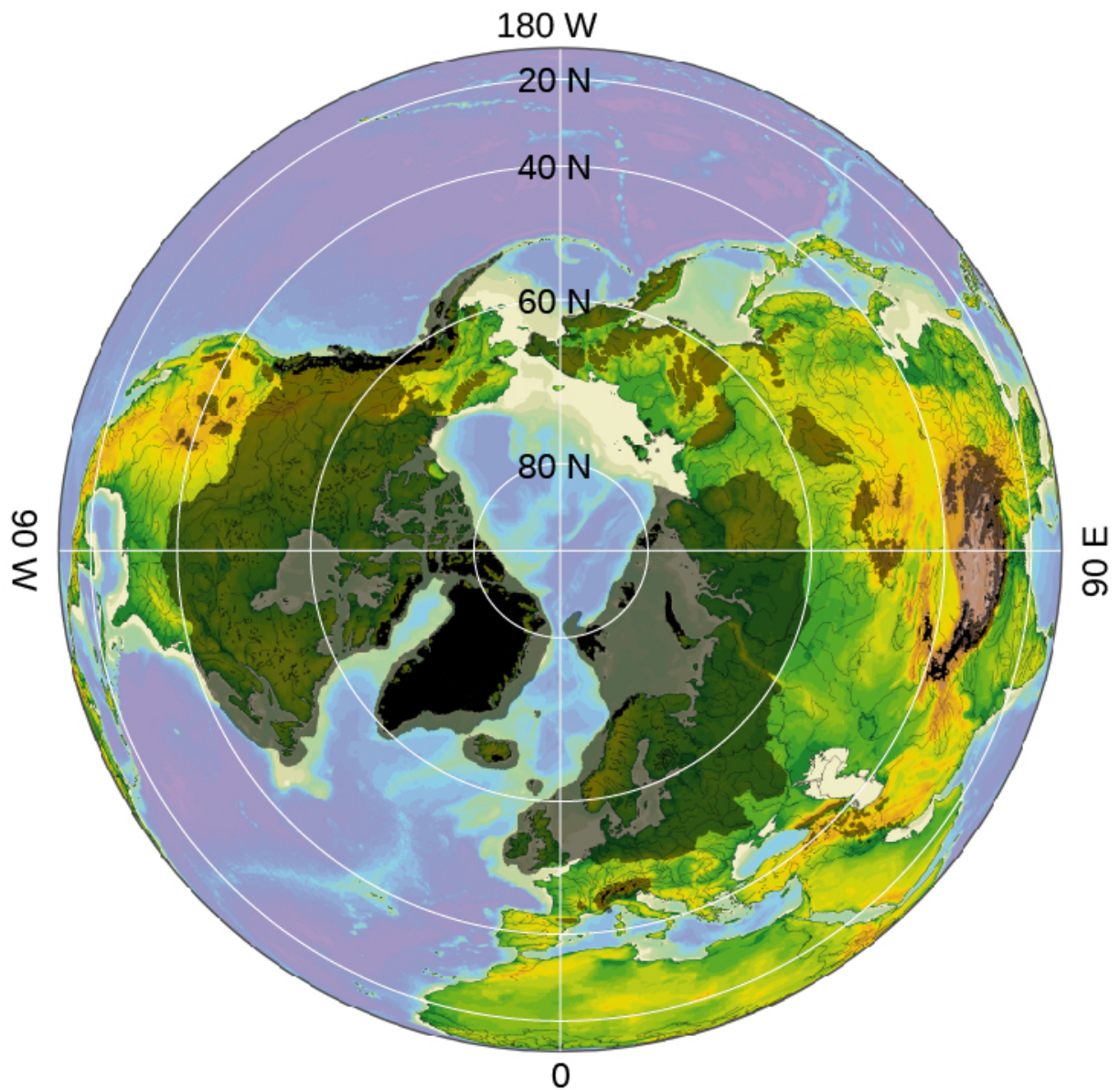
This satellite image shows Hurricane Irene in 2011, shortly before the storm hit land in New York City. The combination of Earth's tilted axis of rotation, moderately rapid rotation, and oceans of liquid water can lead to violent weather on our planet. (credit: NASA/NOAA GOES Project)

Climate is a term used to refer to the effects of the atmosphere that last through decades and centuries. Changes in climate (as opposed to the random variations in weather from one year to the next) are often difficult to detect over short time periods, but as they accumulate, their effect can be devastating. One saying is that "Climate is what you expect, and weather is what you get." Modern farming is especially sensitive to temperature and rainfall; for example, calculations indicate that a drop of only 2 °C throughout the growing season would cut the wheat production by half in Canada and the United States. At the other extreme, an increase of 2 °C in

the average temperature of Earth would be enough to melt many glaciers, including much of the ice cover of Greenland, raising sea level by as much as 10 meters, flooding many coastal cities and ports, and putting small islands completely under water.

The best documented changes in Earth's climate are the great ice ages, which have lowered the temperature of the Northern Hemisphere periodically over the past half million years or so ([\[link\]](#)). The last ice age, which ended about 14,000 years ago, lasted some 20,000 years. At its height, the ice was almost 2 kilometers thick over Boston and stretched as far south as New York City.

Ice Age.



This computer-generated image shows the frozen areas of the Northern Hemisphere during past ice ages from the vantage point of looking down on the North Pole. The area in black indicates the most recent glaciation (coverage by glaciers), and the area in gray shows the maximum level of glaciation ever reached. (credit: modification of work by Hannes Grobe/AWI)

These ice ages were primarily the result of changes in the tilt of Earth's rotational axis, produced by the gravitational effects of the other planets. We are less certain about evidence that at least once (and perhaps twice) about a billion years ago, the entire ocean froze over, a situation called *snowball Earth*.

The development and evolution of life on Earth has also produced changes in the composition and temperature of our planet's atmosphere, as we shall see in the next section.

Note:

Watch this [short excerpt](#) from the National Geographic documentary *Earth: The Biography*. In this segment, Dr. Iain Stewart explains the fluid nature of our atmosphere.

Key Concepts and Summary

The atmosphere has a surface pressure of 1 bar and is composed primarily of N_2 and O_2 , plus such important trace gases as H_2O , CO_2 , and O_3 . Its structure consists of the troposphere, stratosphere, mesosphere, and ionosphere. Changing the composition of the atmosphere also influences the temperature. Atmospheric circulation (weather) is driven by seasonally changing deposition of sunlight. Many longer term climate variations, such as the ice ages, are related to changes in the planet's orbit and axial tilt.

Glossary

bar

a force of 100,000 Newtons acting on a surface area of 1 square meter;
the average pressure of Earth's atmosphere at sea level is 1.013 bars

ozone

(O₃) a heavy molecule of oxygen that contains three atoms rather than the more normal two

stratosphere

the layer of Earth's atmosphere above the troposphere and below the ionosphere

troposphere

the lowest level of Earth's atmosphere, where most weather takes place

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Outline the origins and subsequent diversity of life on Earth
- Explain the ways that life and geological activity have influenced the evolution of the atmosphere
- Describe the causes and effects of the atmospheric greenhouse effect and global warming
- Describe the impact of human activity on our planet's atmosphere and ecology

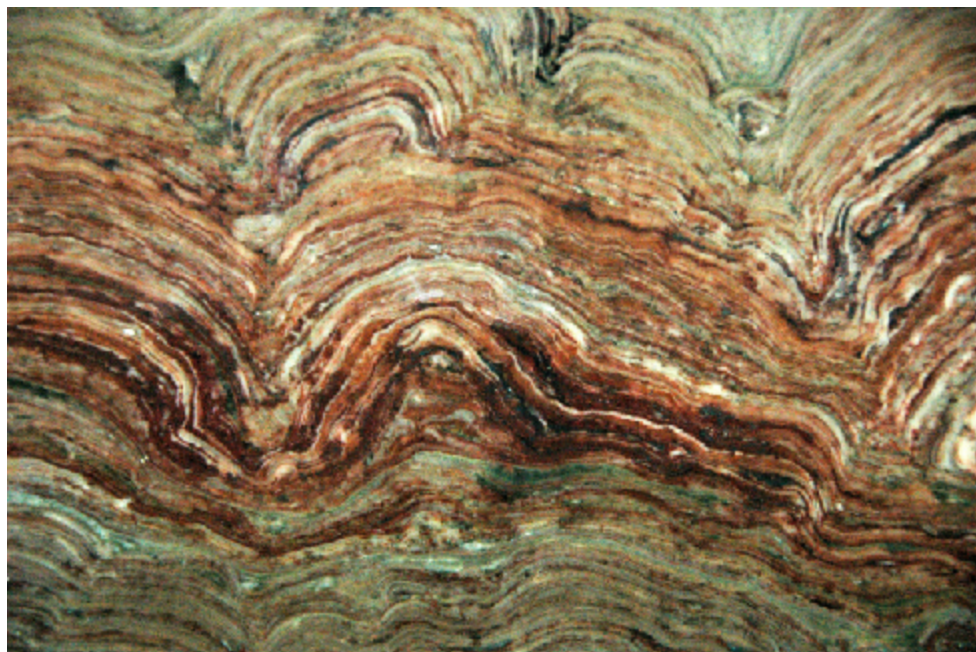
As far as we know, Earth seems to be the only planet in the solar system with life. The origin and development of life are an important part of our planet's story. Life arose early in Earth's history, and living organisms have been interacting with their environment for billions of years. We recognize that life-forms have evolved to adapt to the environment on Earth, and we are now beginning to realize that Earth itself has been changed in important ways by the presence of living matter. The study of the coevolution of life and our planet is one of the subjects of the modern science of *astrobiology*.

The Origin of Life

The record of the birth of life on Earth has been lost in the restless motions of the crust. According to chemical evidence, by the time the oldest surviving rocks were formed about 3.9 billion years ago, life already existed. At 3.5 billion years ago, life had achieved the sophistication to

build large colonies called *stromatolites*, a form so successful that stromatolites still grow on Earth today ([link](#)). But, few rocks survive from these ancient times, and abundant fossils have been preserved only during the past 600 million years—less than 15% of our planet’s history.

Cross-Sections of Fossil Stromatolites.



This polished cross-section of a fossilized colony of stromatolites dates to the Precambrian Era. The layered, domelike structures are mats of sediment trapped in shallow waters by large numbers of blue-green bacteria that can photosynthesize. Such colonies of microorganisms date back more than 3 billion years.

(credit: James St. John)

There is little direct evidence about the actual origin of life. We know that the atmosphere of early Earth, unlike today’s, contained abundant carbon dioxide and some methane, but no oxygen gas. In the absence of oxygen, many complex chemical reactions are possible that lead to the production of amino acids, proteins, and other chemical building blocks of life. Therefore, it seems likely that these chemical building blocks were available very early in Earth’s history and they would have combined to make living organisms.

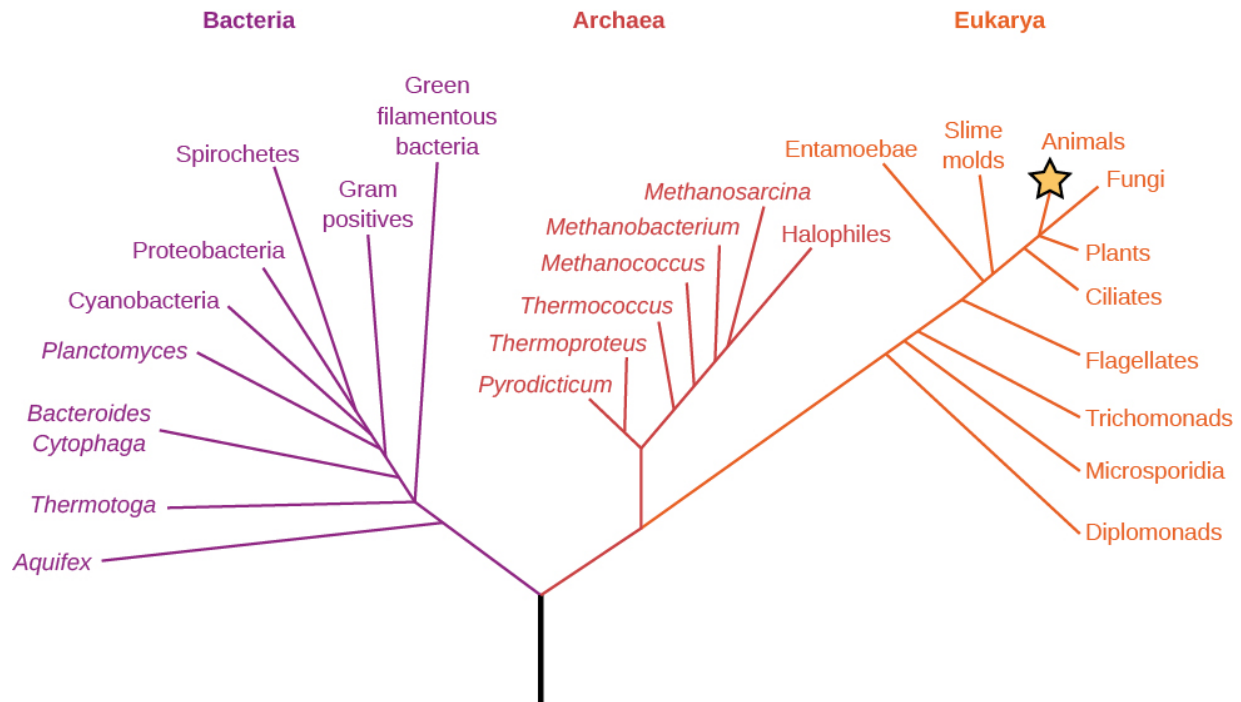
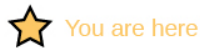
For tens of millions of years after Earth's formation, life (perhaps little more than large molecules, like the viruses of today) probably existed in warm, nutrient-rich seas, living off accumulated organic chemicals. When this easily accessible food became depleted, life began the long evolutionary road that led to the vast numbers of different organisms on Earth today. As it did so, life began to influence the chemical composition of the atmosphere.

In addition to the study of life's history as revealed by chemical and fossil evidence in ancient rocks, scientists use tools from the rapidly advancing fields of genetics and *genomics*—the study of the genetic code that is shared by all life on Earth. While each individual has a unique set of genes (which is why genetic “fingerprinting” is so useful for the study of crime), we also have many genetic traits in common. Your *genome*, the complete map of the DNA in your body, is identical at the 99.9% level to that of Julius Caesar or Marie Curie. At the 99% level, human and chimpanzee genomes are the same. By looking at the gene sequences of many organisms, we can determine that all life on Earth is descended from a common ancestor, and we can use the genetic variations among species as a measure of how closely different species are related.

These genetic analysis tools have allowed scientists to construct what is called the “tree of life” ([\[link\]](#)). This diagram illustrates the way organisms are related by examining one sequence of the nucleic acid RNA that all species have in common. This figure shows that life on Earth is dominated by microscopic creatures that you have probably never heard of. Note that the plant and animal kingdoms are just two little branches at the far right. Most of the diversity of life, and most of our evolution, has taken place at the microbial level. Indeed, it may surprise you to know that there are more microbes in a bucket of soil than there are stars in the Galaxy. You may want to keep this in mind when, later in this book, we turn to the search for life on other worlds. The “aliens” that are most likely to be out there are microbes.

Tree of Life.

Tree of Life



This chart shows the main subdivisions of life on Earth and how they are related. Note that the animal and plant kingdoms are just short branches on the far right, along with the fungi. The most fundamental division of Earth's living things is onto three large domains called bacteria, archaea, and eukarya. Most of the species listed are microscopic. (credit: modification of work by Eric Gaba)

Such genetic studies lead to other interesting conclusions as well. For example, it appears that the earliest surviving terrestrial life-forms were all adapted to live at high temperatures. Some biologists think that life might actually have begun in locations on our planet that were extremely hot. Yet another intriguing possibility is that life began on Mars (which cooled sooner) rather than Earth and was “seeded” onto our planet by meteorites traveling from Mars to Earth. Mars rocks are still making their way to

Earth, but so far none has shown evidence of serving as a “spaceship” to carry microorganisms from Mars to Earth.

The Evolution of the Atmosphere

One of the key steps in the evolution of life on Earth was the development of blue-green algae, a very successful life-form that takes in carbon dioxide from the environment and releases oxygen as a waste product. These successful microorganisms proliferated, giving rise to all the lifeforms we call plants. Since the energy for making new plant material from chemical building blocks comes from sunlight, we call the process **photosynthesis**.

Studies of the chemistry of ancient rocks show that Earth’s atmosphere lacked abundant free oxygen until about 2 billion years ago, despite the presence of plants releasing oxygen by photosynthesis. Apparently, chemical reactions with Earth’s crust removed the oxygen gas as quickly as it formed. Slowly, however, the increasing evolutionary sophistication of life led to a growth in the plant population and thus increased oxygen production. At the same time, it appears that increased geological activity led to heavy erosion on our planet’s surface. This buried much of the plant carbon before it could recombine with oxygen to form CO₂.

Free oxygen began accumulating in the atmosphere about 2 billion years ago, and the increased amount of this gas led to the formation of Earth’s ozone layer (recall that ozone is a triple molecule of oxygen, O₃), which protects the surface from deadly solar ultraviolet light. Before that, it was unthinkable for life to venture outside the protective oceans, so the landmasses of Earth were barren.

The presence of oxygen, and hence ozone, thus allowed colonization of the land. It also made possible a tremendous proliferation of animals, which lived by taking in and using the organic materials produced by plants as their own energy source.

As animals evolved in an environment increasingly rich in oxygen, they were able to develop techniques for breathing oxygen directly from the atmosphere. We humans take it for granted that plenty of free oxygen is

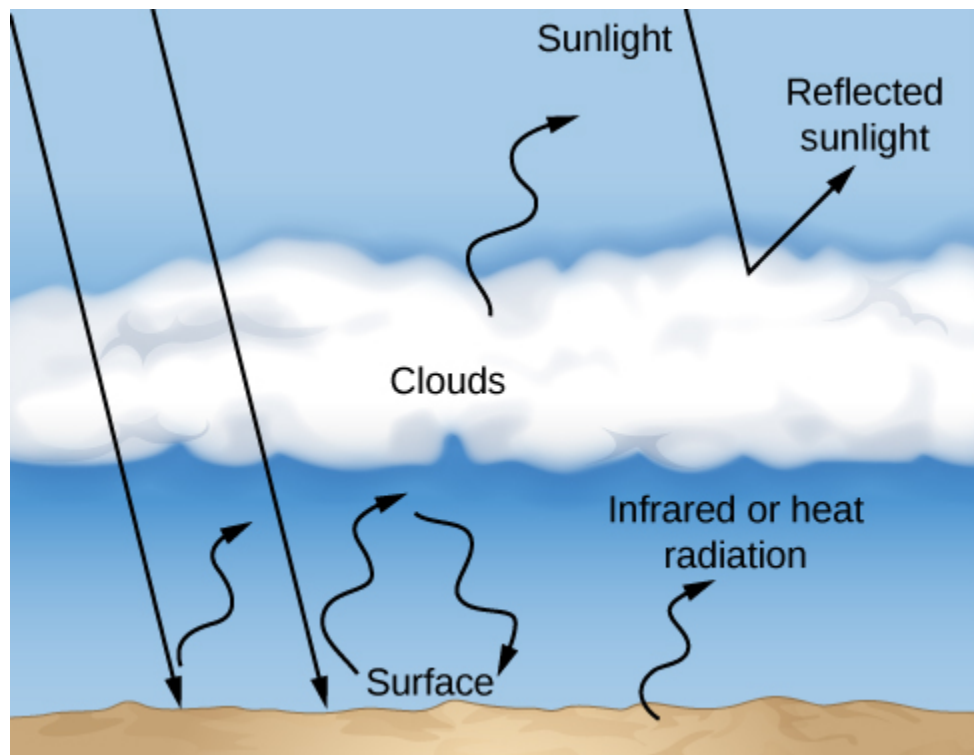
available in Earth's atmosphere, and we use it to release energy from the food we take in. Although it may seem funny to think of it this way, we are lifeforms that have evolved to breathe in the waste product of plants. It is plants and related microbes that are the primary producers, using sunlight to create energy-rich "food" for the rest of us.

On a planetary scale, one of the consequences of life has been a decrease in atmospheric carbon dioxide. In the absence of life, Earth would probably have an atmosphere dominated by CO₂, like Mars or Venus. But living things, in combination with high levels of geological activity, have effectively stripped our atmosphere of most of this gas.

The Greenhouse Effect and Global Warming

We have a special interest in the carbon dioxide content of the atmosphere because of the key role this gas plays in retaining heat from the Sun through a process called the **greenhouse effect**. To understand how the greenhouse effect works, consider the fate of sunlight that strikes the surface of Earth. The light penetrates our atmosphere, is absorbed by the ground, and heats the surface layers. At the temperature of Earth's surface, that energy is then reemitted as infrared or heat radiation ([\[link\]](#)). However, the molecules of our atmosphere, which allow visible light through, are good at absorbing infrared energy. As a result, CO₂ (along with methane and water vapor) acts like a blanket, trapping heat in the atmosphere and impeding its flow back to space. To maintain an energy balance, the temperature of the surface and lower atmosphere must increase until the total energy radiated by Earth to space equals the energy received from the Sun. The more CO₂ there is in our atmosphere, the higher the temperature at which Earth's surface reaches a new balance.

How the Greenhouse Effect Works.



Sunlight that penetrates to Earth's lower atmosphere and surface is reradiated as infrared or heat radiation, which is trapped by greenhouse gases such as water vapor, methane, and CO₂ in the atmosphere. The result is a higher surface temperature for our planet.

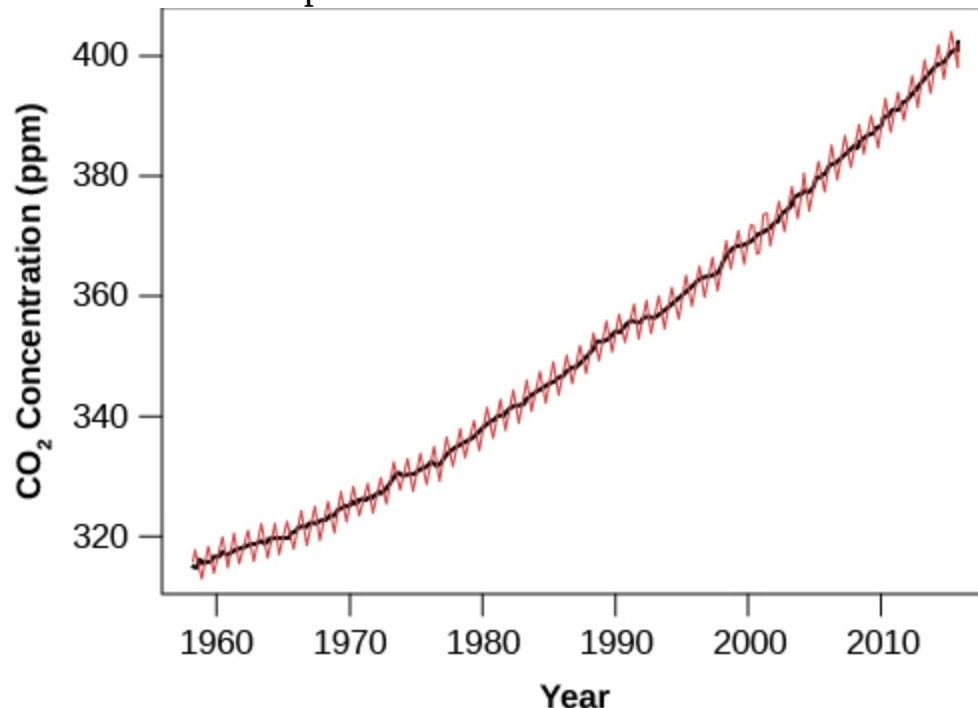
The greenhouse effect in a planetary atmosphere is similar to the heating of a gardener's greenhouse or the inside of a car left out in the Sun with the windows rolled up. In these examples, the window glass plays the role of **greenhouse gases**, letting sunlight in but reducing the outward flow of heat radiation. As a result, a greenhouse or car interior winds up much hotter than would be expected from the heating of sunlight alone. On Earth, the current greenhouse effect elevates the surface temperature by about 23 °C. Without this greenhouse effect, the average surface temperature would be well below freezing and Earth would be locked in a global ice age.

That's the good news; the bad news is that the heating due to the greenhouse effect is increasing. Modern industrial society depends on

energy extracted from burning fossil fuels. In effect, we are exploiting the energy-rich material created by photosynthesis tens of millions of years ago. As these ancient coal and oil deposits are oxidized (burned using oxygen), large quantities of carbon dioxide are released into the atmosphere. The problem is exacerbated by the widespread destruction of tropical forests, which we depend on to extract CO_2 from the atmosphere and replenish our supply of oxygen. In the past century of increased industrial and agricultural development, the amount of CO_2 in the atmosphere increased by about 30% and continues to rise at more than 0.5% per year.

Before the end of the present century, Earth's CO_2 level is predicted to reach twice the value it had before the industrial revolution ([\[link\]](#)). The consequences of such an increase for Earth's surface and atmosphere (and the creatures who live there) are likely to be complex changes in climate, and may be catastrophic for many species. Many groups of scientists are now studying the effects of such global warming with elaborate computer models, and climate change has emerged as the greatest known threat (barring nuclear war) to both industrial civilization and the ecology of our planet.

Increase of Atmospheric Carbon Dioxide over Time.



Scientists expect that the amount of CO₂ will double its preindustrial level before the end of the twenty-first century. Measurements of the isotopic signatures of this added CO₂ demonstrate that it is mostly coming from burning fossil fuels. (credit: modification of work by NOAA)

Note:

This [short PBS video](#) explains the physics of the greenhouse effect.

Already climate change is widely apparent. Around the world, temperature records are constantly set and broken; all but one of the hottest recorded years have taken place since 2000. Glaciers are retreating, and the Arctic Sea ice is now much thinner than when it was first explored with nuclear submarines in the 1950s. Rising sea levels (from both melting glaciers and expansion of the water as its temperature rises) pose one of the most immediate threats, and many coastal cities have plans to build dikes or seawalls to hold back the expected flooding. The rate of temperature increase is without historical precedent, and we are rapidly entering “unknown territory” where human activities are leading to the highest temperatures on Earth in more than 50 million years.

Human Impacts on Our Planet

Earth is so large and has been here for so long that some people have trouble accepting that humans are really changing the planet, its atmosphere, and its climate. They are surprised to learn, for example, that the carbon dioxide released from burning fossil fuels is 100 times greater than that emitted by volcanoes. But, the data clearly tell the story that our climate is changing rapidly, and that almost all of the change is a result of human activity.

This is not the first time that humans have altered our environment dramatically. Some of the greatest changes were caused by our ancestors, before the development of modern industrial society. If aliens had visited Earth 50,000 years ago, they would have seen much of the planet supporting large animals of the sort that now survive only in Africa. The plains of Australia were occupied by giant marsupials such as diprododon and zygomaturus (the size of our elephants today), and a species of kangaroo that stood 10 feet high. North America and North Asia hosted mammoths, saber tooth cats, mastodons, giant sloths, and even camels. The Islands of the Pacific teemed with large birds, and vast forests covered what are now the farms of Europe and China. Early human hunters killed many large mammals and marsupials, early farmers cut down most of the forests, and the Polynesian expansion across the Pacific doomed the population of large birds.

An even greater mass extinction is underway as a result of rapid climate change. In recognition of our impact on the environment, scientists have proposed giving a new name to the current epoch, the *anthropocene*, when human activity started to have a significant global impact. Although not an officially approved name, the concept of “anthropocene” is useful for recognizing that we humans now represent the dominant influence on our planet’s atmosphere and ecology, for better or for worse.

Key Concepts and Summary

Life originated on Earth at a time when the atmosphere lacked O₂ and consisted mostly of CO₂. Later, photosynthesis gave rise to free oxygen and ozone. Modern genomic analysis lets us see how the wide diversity of species on the planet are related to each other. CO₂ and methane in the atmosphere heat the surface through the greenhouse effect; today, increasing amounts of atmospheric CO₂ are leading to the global warming of our planet.

Glossary

greenhouse gas

a gas in an atmosphere that absorbs and emits radiation within the thermal infrared range; on Earth, these atmospheric gases primarily include carbon dioxide, methane, and water vapor

greenhouse effect

the blanketing (absorption) of infrared radiation near the surface of a planet—for example, by CO₂ in its atmosphere

photosynthesis

a complex sequence of chemical reactions through which some living things can use sunlight to manufacture products that store energy (such as carbohydrates), releasing oxygen as one by-product

The Geology of Venus

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the general features of the surface of Venus
- Explain what the study of craters on Venus tells us about the age of its surface
- Compare tectonic activity and volcanoes on Venus with those of Earth
- Explain why the surface of Venus is inhospitable to human life

Since Venus has about the same size and composition as Earth, we might expect its geology to be similar. This is partly true, but Venus does not exhibit the same kind of *plate tectonics* as Earth, and we will see that its lack of erosion results in a very different surface appearance.

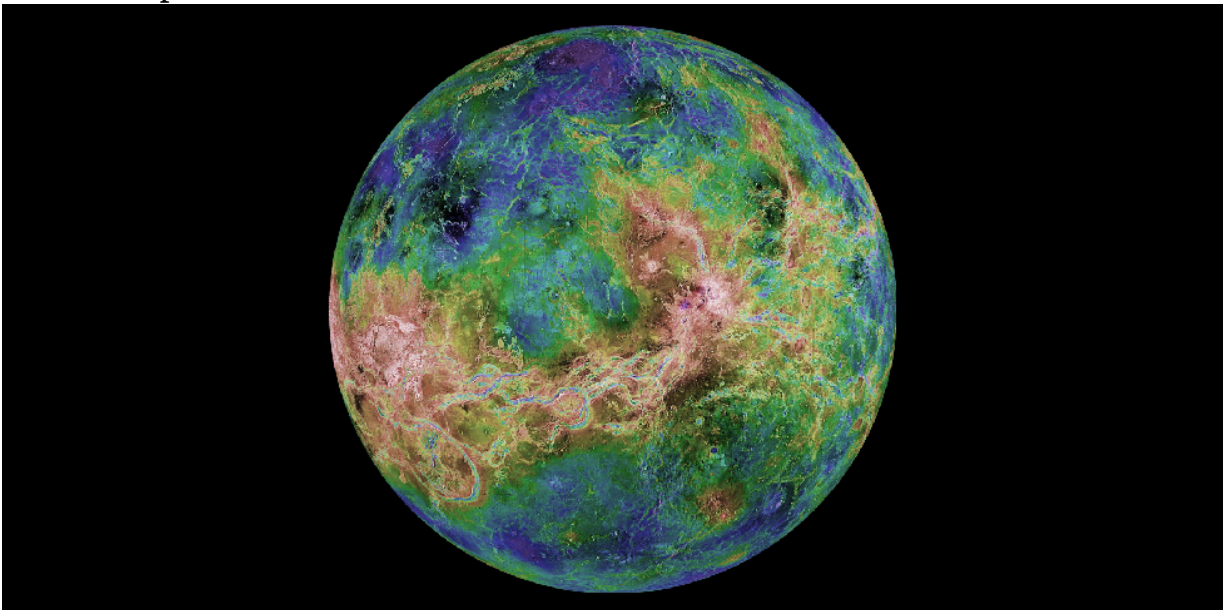
Spacecraft Exploration of Venus

Nearly 50 spacecraft have been launched to Venus, but only about half were successful. Although the 1962 US Mariner 2 flyby was the first, the Soviet Union launched most of the subsequent missions to Venus. In 1970, Venera 7 became the first probe to land and broadcast data from the surface of Venus. It operated for 23 minutes before succumbing to the high surface temperature. Additional Venera probes and landers followed, photographing the surface and analyzing the atmosphere and soil.

To understand the geology of Venus, however, we needed to make a global study of its surface, a task made very difficult by the perpetual cloud layers surrounding the planet. The problem resembles the challenge facing air traffic controllers at an airport, when the weather is so cloudy or smoggy that they can't locate the incoming planes visually. The solution is similar in both cases: use a radar instrument to probe through the obscuring layer.

The first global radar map was made by the US Pioneer Venus orbiter in the late 1970s, followed by better maps from the twin Soviet Venera 15 and 16 radar orbiters in the early 1980s. However, most of our information on the geology of Venus is derived from the US *Magellan* spacecraft, which mapped Venus with a powerful *imaging radar*. *Magellan* produced images with a resolution of 100 meters, much better than that of previous missions, yielding our first detailed look at the surface of our sister planet ([\[link\]](#)). (The *Magellan* spacecraft returned more data to Earth than all previous planetary missions combined; each 100 minutes of data transmission from the spacecraft provided enough information, if translated into characters, to fill two 30-volume encyclopedias.)

Radar Map of Venus.



This composite image has a resolution of about 3 kilometers. Colors have been added to indicate elevation, with blue meaning low and brown and white high. The large continent Aphrodite stretches around the equator, where the bright (therefore rough) surface has been

deformed by tectonic forces in the crust of Venus. (credit: modification of work by NASA/JPL/USGS)

Consider for a moment how good *Magellan's* resolution of 100 meters really is. It means the radar images from Venus can show anything on the surface larger than a football field. Suddenly, a whole host of topographic features on Venus became accessible to our view. As you look at the radar images throughout this chapter, bear in mind that these are constructed from radar reflections, not from visible-light photographs. For example, bright features on these radar images are an indication of rough terrain, whereas darker regions are smoother.

Probing Through the Clouds of Venus

The radar maps of Venus reveal a planet that looks much the way Earth might look if our planet's surface were not constantly being changed by erosion and deposition of sediment. Because there is no water or ice on Venus and the surface wind speeds are low, almost nothing obscures or erases the complex geological features produced by the movements of Venus' crust, by volcanic eruptions, and by impact craters. Having finally penetrated below the clouds of Venus, we find its surface to be naked, revealing the history of hundreds of millions of years of geological activity.

About 75% of the surface of Venus consists of lowland lava plains. Superficially, these plains resemble the basaltic ocean basins of Earth, but they were not produced in quite the same way. There is no evidence of subduction zones on Venus, indicating that, unlike Earth, this planet never experienced plate tectonics. Although *convection* (the rising of hot materials) in its mantle generated great stresses in the crust of Venus, they did not start large continental plates moving. The formation of the lava plains of Venus more nearly resembles that of the lunar maria. Both were the result of widespread lava eruptions without the crustal spreading associated with plate tectonics.

Rising above the lowland lava plains are two full-scale continents of mountainous terrain. The largest continent on Venus, called Aphrodite, is about the size of Africa (you can see it stand out in [\[link\]](#)). Aphrodite stretches along the equator for about one-third of the way around the planet. Next in size is the northern highland region Ishtar, which is about the size of Australia. Ishtar contains the highest region on the planet, the Maxwell Mountains, which rise 11 kilometers above the surrounding lowlands. (The Maxwell Mountains are the only feature on Venus named after a man. They commemorate James Clerk Maxwell, whose theory of electromagnetism led to the invention of radar. All other features are named for women, either from history or mythology.)

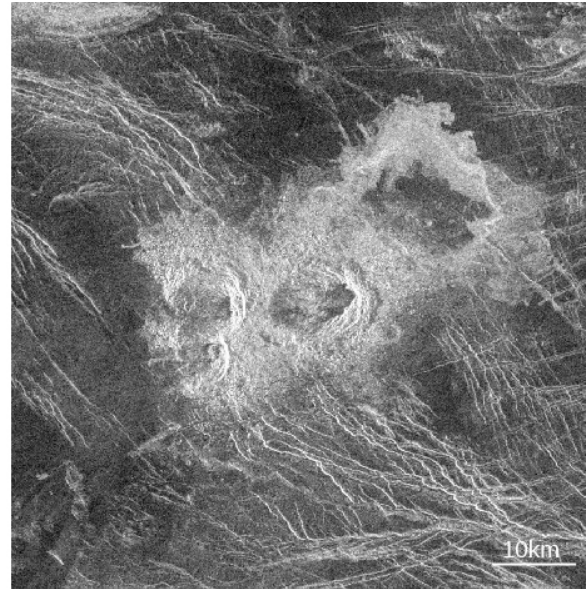
Craters and the Age of the Venus Surface

One of the first questions astronomers addressed with the high-resolution *Magellan* images was the age of the surface of Venus. Remember that the age of a planetary surface is rarely the age of the world it is on. A young age merely implies an active geology in that location. Such ages can be derived from counting impact craters. [\[link\]](#) is an example of what these craters look like on the Venus radar images. The more densely cratered the surface, the greater its age. The largest crater on Venus (called Mead) is 275 kilometers in diameter, slightly larger than the largest known terrestrial crater (Chicxulub), but much smaller than the lunar impact basins.

Impact Craters on Venus.



(a)



(b)

(a) These large impact craters are in the Lavinia region of Venus. Because they are rough, the crater rims and ejecta appear brighter in these radar images than do the smoother surrounding lava plains. The largest of these craters has a diameter of 50 kilometers. (b) This small, complex crater is named after writer Gertrude Stein. The triple impact was caused by the breaking apart of the incoming asteroid during its passage through the thick atmosphere of Venus. The projectile had an initial diameter of between 1 and 2 kilometers. (credit a: modification of work by NASA/JPL; credit b: modification of work by NASA/JPL)

You might think that the thick atmosphere of Venus would protect the surface from impacts, burning up the projectiles long before they could reach the surface. But this is the case for only smaller projectiles. Crater statistics show very few craters less than 10 kilometers in diameter, indicating that projectiles smaller than about 1 kilometer (the size that typically produces a 10-kilometer crater) were stopped by the atmosphere. Those craters with diameters from 10 to 30 kilometers are frequently distorted or multiple, apparently because the incoming projectile broke apart in the atmosphere before it could strike the ground as shown in the Stein crater in [\[link\]](#). If we limit ourselves to impacts that produce craters

with diameters of 30 kilometers or larger, however, then crater counts are as useful on Venus for measuring surface age as they are on airless bodies such as the Moon.

The large craters in the venusian plains indicate an average surface age that is only between 300 and 600 million years. These results indicate that Venus is indeed a planet with persistent geological activity, intermediate between that of Earth's ocean basins (which are younger and more active) and that of its continents (which are older and less active).

Almost all of the large craters on Venus look fresh, with little degradation or filling in by either lava or windblown dust. This is one way we know that the rates of erosion or sediment deposition are very low. We have the impression that relatively little has happened since the venusian plains were last resurfaced by large-scale volcanic activity. Apparently Venus experienced some sort of planet-wide volcanic convulsion between 300 and 600 million years ago, a mysterious event that is unlike anything in terrestrial history.

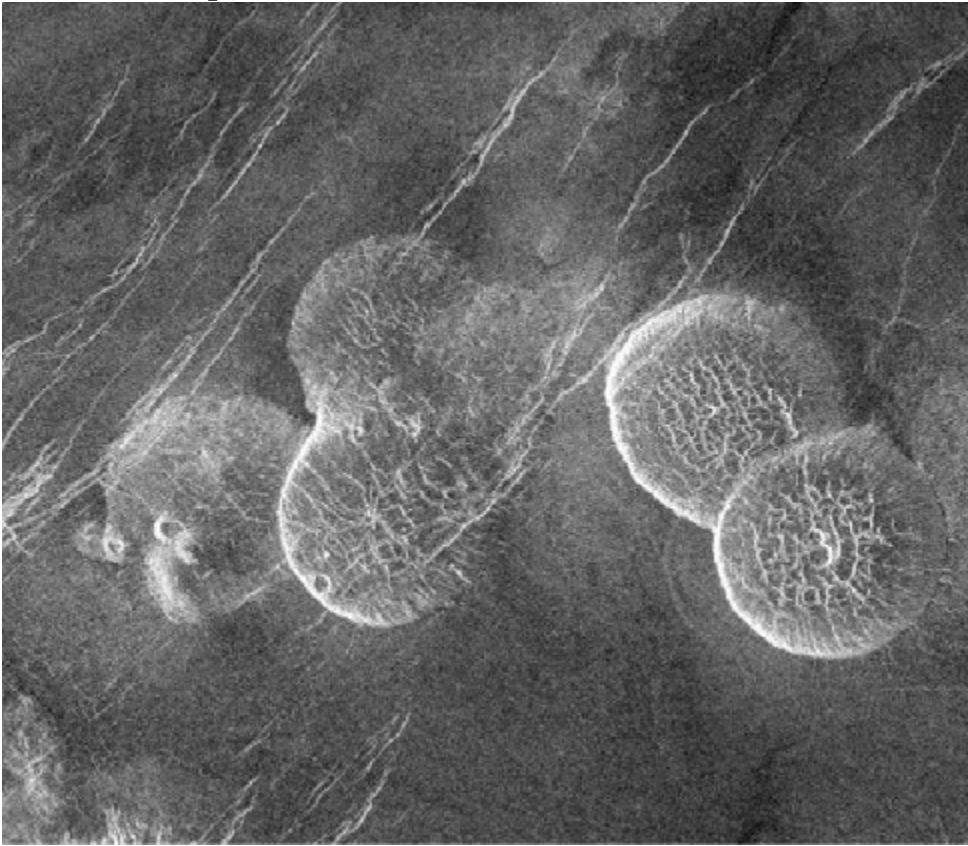
Volcanoes on Venus

Like Earth, Venus is a planet that has experienced widespread volcanism. In the lowland plains, volcanic eruptions are the principal way the surface is renewed, with large flows of highly fluid lava destroying old craters and generating a fresh surface. In addition, numerous younger volcanic mountains and other structures are associated with surface hot spots—places where convection in the planet's mantle transports the interior heat to the surface.

The largest individual volcano on Venus, called Sif Mons, is about 500 kilometers across and 3 kilometers high—broader but lower than the Hawaiian volcano Mauna Loa. At its top is a volcanic crater, or *caldera*, about 40 kilometers across, and its slopes show individual lava flows up to 500 kilometers long. Thousands of smaller volcanoes dot the surface, down to the limit of visibility of the *Magellan* images, which correspond to cones or domes about the size of a shopping mall parking lot. Most of these seem

similar to terrestrial volcanoes. Other volcanoes have unusual shapes, such as the “pancake domes” illustrated in [\[link\]](#).

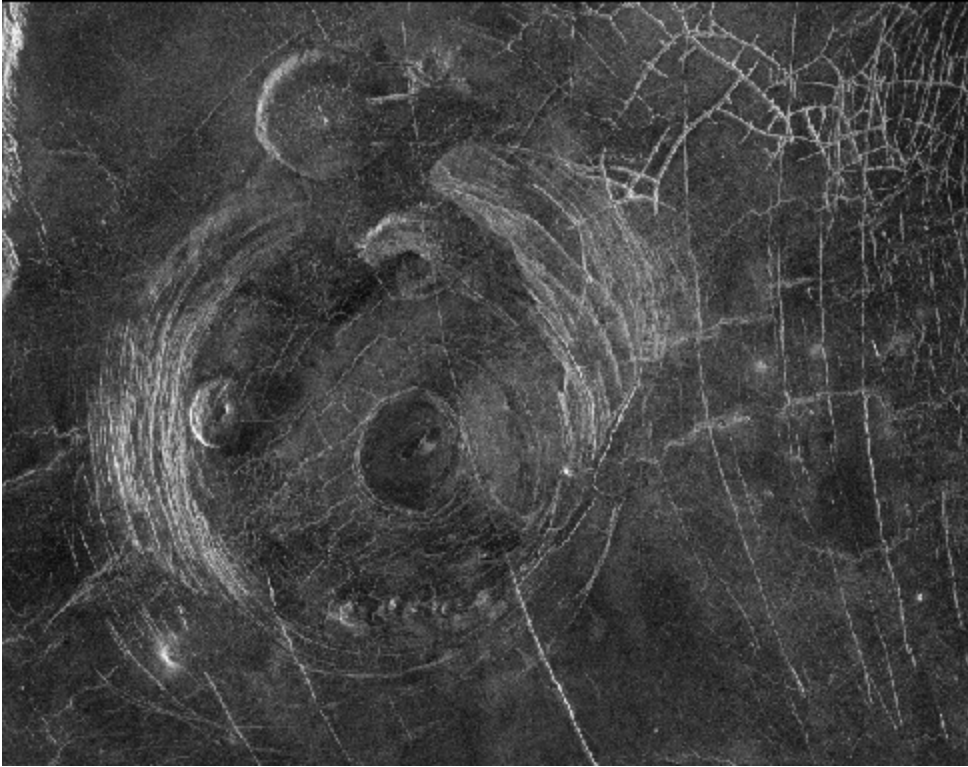
Pancake-Shaped Volcanoes on Venus.



These remarkable circular domes, each about 25 kilometers across and about 2 kilometers tall, are the result of eruptions of highly viscous (sludgy) lava that spreads out evenly in all directions. (credit: modification of work by NASA/JPL)

All of the volcanism is the result of eruption of lava onto the surface of the planet. But the hot lava rising from the interior of a planet does not always make it to the surface. On both Earth and Venus, this upwelling lava can collect to produce bulges in the crust. Many of the granite mountain ranges on Earth, such as the Sierra Nevada in California, involve such subsurface volcanism. These bulges are common on Venus, where they produce large circular or oval features called *coronae* (singular: corona) ([\[link\]](#)).

The “Miss Piggy” Corona.



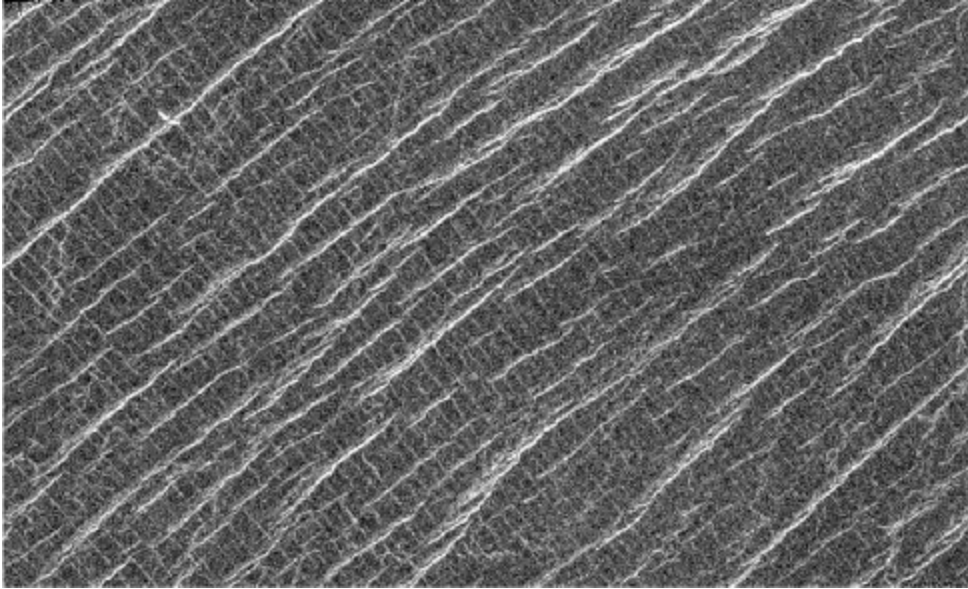
Fotla Corona is located in the plains to the south of Aphrodite Terra. Curved fracture patterns show where the material beneath has put stress on the surface. A number of pancake and dome volcanoes are also visible. Fotla was a Celtic fertility goddess. Some students see a resemblance between this corona and Miss Piggy of the Muppets (her left ear, at the top of the picture, is the pancake volcano in the upper center of the image). (credit: NASA/JPL)

Tectonic Activity

Convection currents of molten material in the mantle of Venus push and stretch the crust. Such forces are called **tectonic**, and the geological features that result from these forces are called *tectonic features*. On Venus' lowland plains, tectonic forces have broken the lava surface to create remarkable

patterns of ridges and cracks ([link](#)). In a few places, the crust has even torn apart to generate rift valleys. The circular features associated with coronae are tectonic ridges and cracks, and most of the mountains of Venus also owe their existence to tectonic forces.

Ridges and Cracks.



This region of the Lakshmi Plains on Venus has been fractured by tectonic forces to produce a cross-hatched grid of cracks and ridges. Be sure to notice the fainter linear features that run perpendicular to the brighter ones. As this is a radar image, the brightness of the ridges indicates their relative height. This image shows a region about 80 kilometers wide and 37 kilometers high. Lakshmi is a Hindu goddess of prosperity. (credit: modification of work by Magellan Team, JPL, NASA)

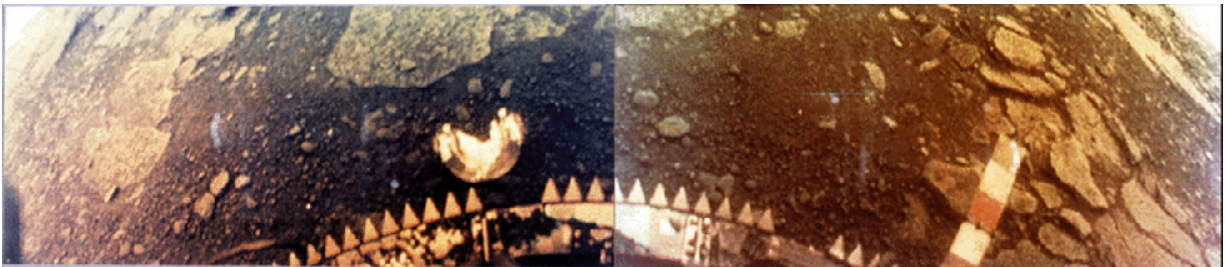
The Ishtar continent, which has the highest elevations on Venus, is the most dramatic product of these tectonic forces. Ishtar and its tall Maxwell Mountains resemble the Tibetan Plateau and Himalayan Mountains on Earth. Both are the product of compression of the crust, and both are maintained by the continuing forces of mantle convection.

On Venus' Surface

The successful Venera landers of the 1970s found themselves on an extraordinarily inhospitable planet, with a surface pressure of 90 bars and a temperature hot enough to melt lead and zinc. Despite these unpleasant conditions, the spacecraft were able to photograph their surroundings and collect surface samples for chemical analysis before their instruments gave out. The diffuse sunlight striking the surface was tinted red by the clouds, and the illumination level was equivalent to a heavy overcast on Earth.

The probes found that the rock in the landing areas is igneous, primarily basalts. Examples of the Venera photographs are shown in [\[link\]](#). Each picture shows a flat, desolate landscape with a variety of rocks, some of which may be ejecta from impacts. Other areas show flat, layered lava flows. There have been no further landings on Venus since the 1970s.

Surface of Venus.



These views of the surface of Venus are from the Venera 13 spacecraft. Everything is orange because the thick atmosphere of Venus absorbs the bluer colors of light. The horizon is visible in the upper corner of each image. (credit: NASA)

Key Concepts and Summary

Venus has been mapped by radar, especially with the *Magellan* spacecraft. Its crust consists of 75% lowland lava plains, numerous volcanic features, and many large coronae, which are the expression of subsurface volcanism. The planet has been modified by widespread tectonics driven by mantle convection, forming complex patterns of ridges and cracks and building

high continental regions such as Ishtar. The surface is extraordinarily inhospitable, with pressure of 90 bars and temperature of 730 K, but several Russian Venera landers investigated it successfully.

Glossary

tectonic

geological features that result from stresses and pressures in the crust of a planet; tectonic forces can lead to earthquakes and motion of the crust

Atmospheres of the Giant Planets

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Discuss the atmospheric composition of the giant planets
- Describe the cloud formation and atmospheric structure of the gas giants
- Characterize the giant planets' wind and weather patterns
- Understand the scale and longevity of storms on the giant planets

The atmospheres of the jovian planets are the parts we can observe or measure directly. Since these planets have no solid surfaces, their atmospheres are more representative of their general compositions than is the case with the terrestrial planets. These atmospheres also present us with some of the most dramatic examples of weather patterns in the solar system. As we will see, storms on these planets can grow bigger than the entire planet Earth.

Atmospheric Composition

When sunlight reflects from the atmospheres of the giant planets, the atmospheric gases leave their “fingerprints” in the spectrum of light. Spectroscopic observations of the jovian planets began in the nineteenth century, but for a long time, astronomers were not able to interpret the spectra they observed. As late as the 1930s, the most prominent features photographed in these spectra remained unidentified. Then better spectra

revealed the presence of molecules of methane (CH_4) and ammonia (NH_3) in the atmospheres of Jupiter and Saturn.

At first astronomers thought that methane and ammonia might be the main constituents of these atmospheres, but now we know that hydrogen and helium are actually the dominant gases. The confusion arose because neither hydrogen nor helium possesses easily detected spectral features in the visible spectrum. It was not until the Voyager spacecraft measured the far-infrared spectra of Jupiter and Saturn that a reliable abundance for the elusive helium could be found.

The compositions of the two atmospheres are generally similar, except that on Saturn there is less helium as the result of the precipitation of helium that contributes to Saturn's internal energy source. The most precise measurements of composition were made on Jupiter by the Galileo entry probe in 1995; as a result, we know the abundances of some elements in the jovian atmosphere even better than we know those in the Sun.

Note:

James Van Allen: Several Planets under His Belt

The career of physicist James Van Allen spanned the birth and growth of the space age, and he played a major role in its development. Born in Iowa in 1914, Van Allen received his PhD from the University of Iowa. He then worked for several research institutions and served in the Navy during World War II.

After the war, Van Allen ([\[link\]](#)) was appointed Professor of Physics at the University of Iowa. He and his collaborators began using rockets to explore cosmic radiation in Earth's outer atmosphere. To reach extremely high altitudes, Van Allen designed a technique in which a balloon lifts and then launches a small rocket (the rocket is nicknamed "the rockoon").

James Van Allen (1914–2006).



In this 1950s photograph, Van Allen holds a “rockoon.” (credit: modification of work by Frederick W. Kent Collection, University of Iowa Archives)

Over dinner one night in 1950, Van Allen and several colleagues came up with the idea of the International Geophysical Year (IGY), an opportunity for scientists around the world to coordinate their investigations of the physics of Earth, especially research done at high altitudes. In 1955, the United States and the Soviet Union each committed themselves to launching an Earth-orbiting satellite during IGY, a competition that began what came to be known as the space race. The IGY (stretched to 18 months) took place between July 1957 and December 1958.

The Soviet Union won the first lap of the race by launching Sputnik 1 in October 1957. The US government spurred its scientists and engineers to even greater efforts to get something into space to maintain the country's

prestige. However, the primary US satellite program, Vanguard, ran into difficulties: each of its early launches crashed or exploded. Simultaneously, a second team of rocket engineers and scientists had quietly been working on a military launch vehicle called Jupiter-C. Van Allen spearheaded the design of the instruments aboard a small satellite that this vehicle would carry. On January 31, 1958, Van Allen's Explorer 1 became the first US satellite in space.

Unlike Sputnik, Explorer 1 was equipped to make scientific measurements of high-energy charged particles above the atmosphere. Van Allen and his team discovered a belt of highly charged particles surrounding Earth, and these belts now bear his name. This first scientific discovery of the space program made Van Allen's name known around the world.

Van Allen and his colleagues continued to measure the magnetic and particle environment around planets with increasingly sophisticated spacecraft, including Pioneers 10 and 11, which made exploratory surveys of the environments of Jupiter and Saturn. Some scientists refer to the charged-particle zones around those planets as Van Allen belts as well.

(Once, when Van Allen was giving a lecture at the University of Arizona, the graduate students in planetary science asked him if he would leave his belt at the school. It is now proudly displayed as the university's "Van Allen belt.")

Van Allen was a strong supporter of space science and an eloquent senior spokesperson for the American scientific community, warning NASA not to put all its efforts into human spaceflight, but to also use robotic spacecraft as productive tools for space exploration.

Clouds and Atmospheric Structure

The clouds of Jupiter ([\[link\]](#)) are among the most spectacular sights in the solar system, much beloved by makers of science-fiction films. They range in color from white to orange to red to brown, swirling and twisting in a constantly changing kaleidoscope of patterns. Saturn shows similar but much more subdued cloud activity; instead of vivid colors, its clouds have a nearly uniform butterscotch hue ([\[link\]](#)).

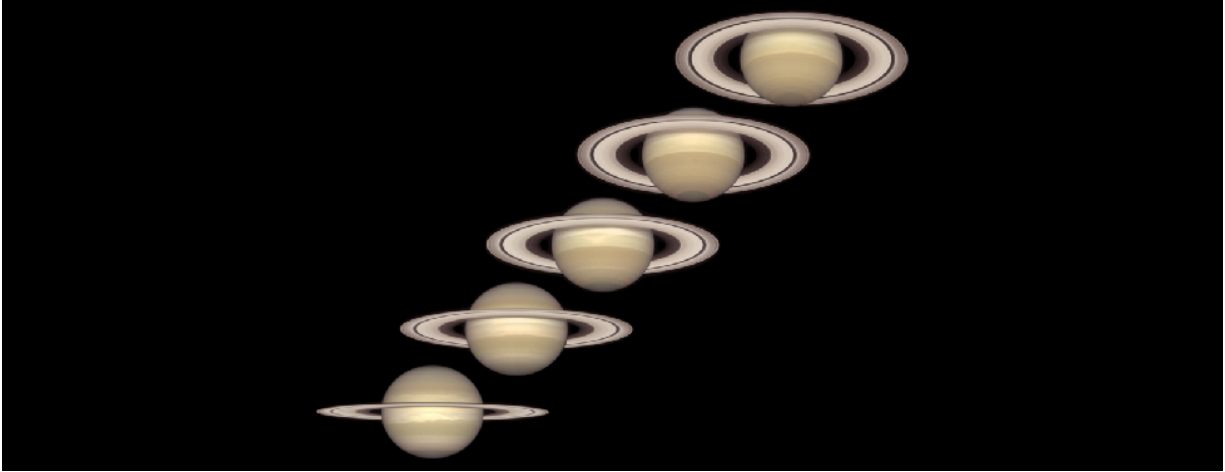
Jupiter's Colorful Clouds.



The vibrant colors of the clouds on Jupiter present a puzzle to astronomers: given the cool temperatures and the composition of nearly 90% hydrogen, the atmosphere should be colorless. One hypothesis suggests that perhaps colorful hydrogen compounds rise from warm areas. The actual colors are a bit more muted, as shown in [\[link\]](#). (credit: modification of work by Voyager Project, JPL, and NASA)

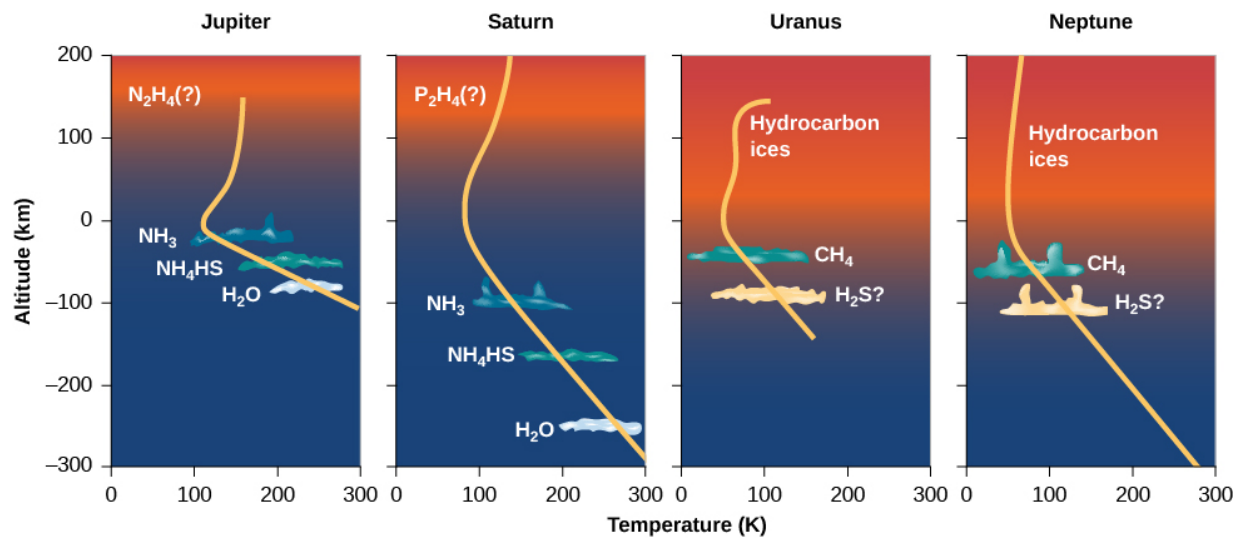
Different gases freeze at different temperatures. At the temperatures and pressures of the upper atmospheres of Jupiter and Saturn, methane remains a gas, but ammonia can condense and freeze. (Similarly, water vapor condenses high in Earth's atmosphere to produce clouds of ice crystals.) The primary clouds that we see around these planets, whether from a

spacecraft or through a telescope, are composed of frozen ammonia crystals. The ammonia clouds mark the upper edge of the planets' tropospheres; above that is the stratosphere, the coldest part of the atmosphere. (These layers were initially defined in [Earth as a Planet](#).)
Saturn over Five Years.



These beautiful images of Saturn were recorded by the Hubble Space Telescope between 1996 and 2000. Since Saturn is tilted by 27° , we see the orientation of Saturn's rings around its equator change as the planet moves along its orbit. Note the horizontal bands in the atmosphere. (credit: modification of work by NASA and The Hubble Heritage Team (STScI/AURA))

The diagrams in [\[link\]](#) show the structure and clouds in the atmospheres of all four jovian planets. On both Jupiter and Saturn, the temperature near the cloud tops is about 140 K (only a little cooler than the polar caps of Mars). On Jupiter, this cloud level is at a pressure of about 0.1 bar (one tenth the atmospheric pressure at the surface of Earth), but on Saturn it occurs lower in the atmosphere, at about 1 bar. Because the ammonia clouds lie so much deeper on Saturn, they are more difficult to see, and the overall appearance of the planet is much blander than is Jupiter's appearance.
Atmospheric Structure of the Jovian Planets.



In each diagram, the yellow line shows how the temperature (see the scale on the bottom) changes with altitude (see the scale at the left). The location of the main layers on each planet is also shown.

Within the tropospheres of these planets, the temperature and pressure both increase with depth. Through breaks in the ammonia clouds, we can see tantalizing glimpses of other cloud layers that can form in these deeper regions of the atmosphere—regions that were sampled directly for Jupiter by the Galileo probe that fell into the planet.

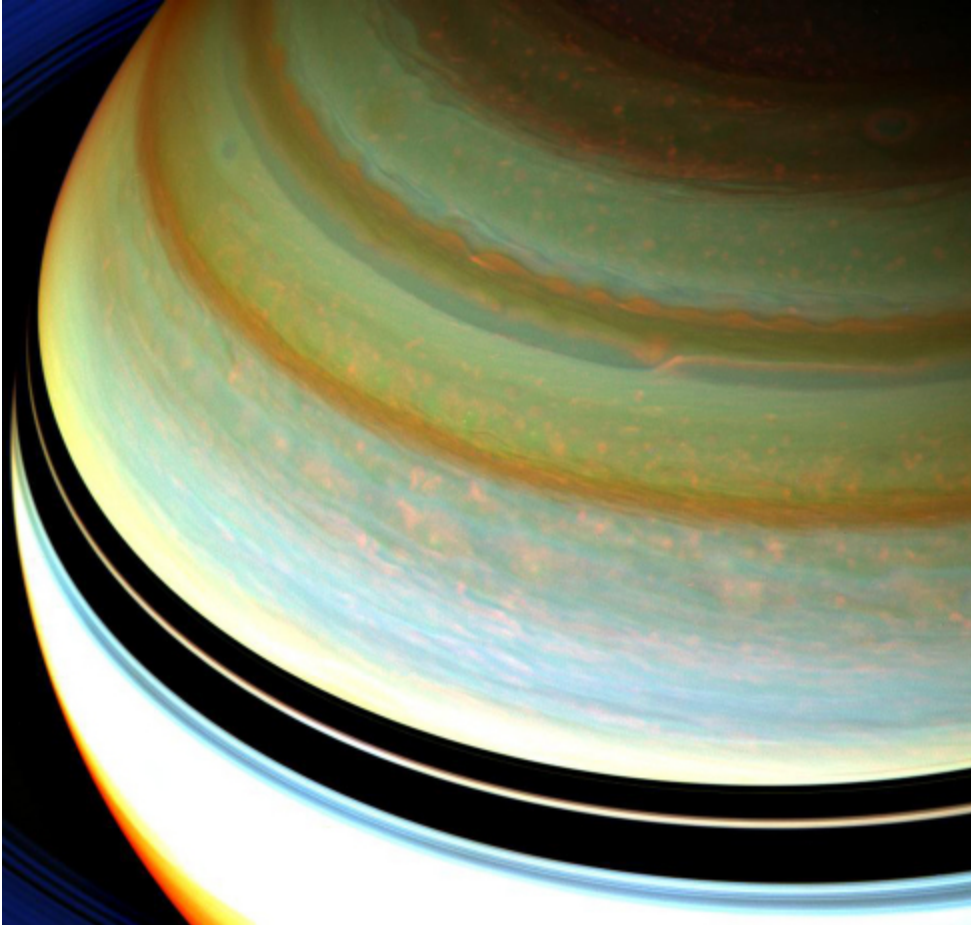
As it descended to a pressure of 5 bars, the probe should have passed into a region of frozen water clouds, then below that into clouds of liquid water droplets, perhaps similar to the common clouds of the terrestrial troposphere. At least this is what scientists expected. But the probe saw no water clouds, and it measured a surprisingly low abundance of water vapor in the atmosphere. It soon became clear to the Galileo scientists that the probe happened to descend through an unusually dry, cloud-free region of the atmosphere—a giant downdraft of cool, dry gas. Andrew Ingersoll of Caltech, a member of the Galileo team, called this entry site the “desert” of Jupiter. It’s a pity that the probe did not enter a more representative region, but that’s the luck of the cosmic draw. The probe continued to make measurements to a pressure of 22 bars but found no other cloud layers before its instruments stopped working. It also detected lightning storms,

but only at great distances, further suggesting that the probe itself was in a region of clear weather.

Above the visible ammonia clouds in Jupiter's atmosphere, we find the clear stratosphere, which reaches a minimum temperature near 120 K. At still higher altitudes, temperatures rise again, just as they do in the upper atmosphere of Earth, because here the molecules absorb ultraviolet light from the Sun. The cloud colors are due to impurities, the product of chemical reactions among the atmospheric gases in a process we call **photochemistry**. In Jupiter's upper atmosphere, photochemical reactions create a variety of fairly complex compounds of hydrogen and carbon that form a thin layer of smog far above the visible clouds. We show this smog as a fuzzy orange region in [\[link\]](#); however, this thin layer does not block our view of the clouds beneath it.

The visible atmosphere of Saturn is composed of approximately 75% hydrogen and 25% helium, with trace amounts of methane, ethane, propane, and other hydrocarbons. The overall structure is similar to that of Jupiter. Temperatures are somewhat colder, however, and the atmosphere is more extended due to Saturn's lower surface gravity. Thus, the layers are stretched out over a longer distance, as you can see in [\[link\]](#). Overall, though, the same atmospheric regions, condensation cloud, and photochemical reactions that we see on Jupiter should be present on Saturn ([\[link\]](#)).

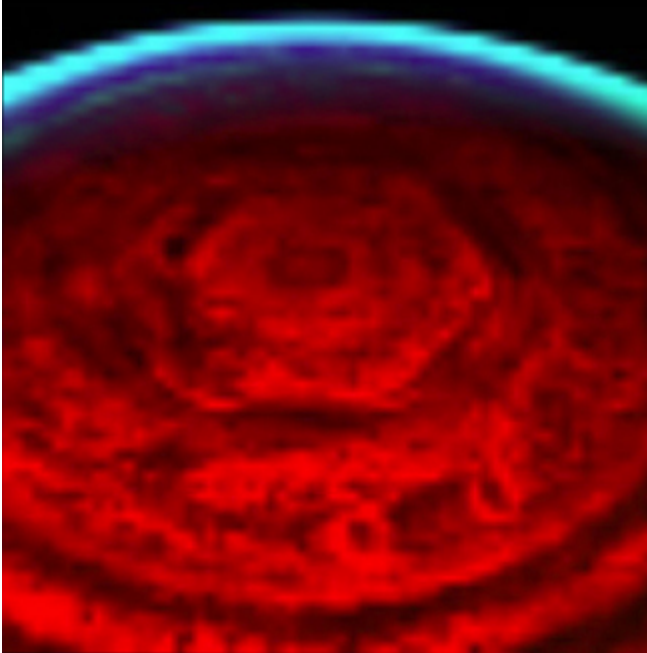
Cloud Structure on Saturn.



In this Cassini image, colors have been intensified, so we can see the bands and zones and storms in the atmosphere. The dark band is the shadow of the rings on the planet. (credit: NASA/JPL-Caltech/Space Science Institute)

Saturn has one anomalous cloud structure that has mystified scientists: a hexagonal wave pattern around the north pole, shown in [\[link\]](#). The six sides of the hexagon are each longer than the diameter of Earth. Winds are also extremely high on Saturn, with speeds of up to 1800 kilometers per hour measured near the equator.

Hexagon Pattern on Saturn's North Pole.



In this infrared nighttime image from the Cassini mission, the path of Saturn's hexagonal jet stream is visible as the planet's north pole emerges from the darkness of winter. (credit: NASA/JPL/University of Arizona)

Note:

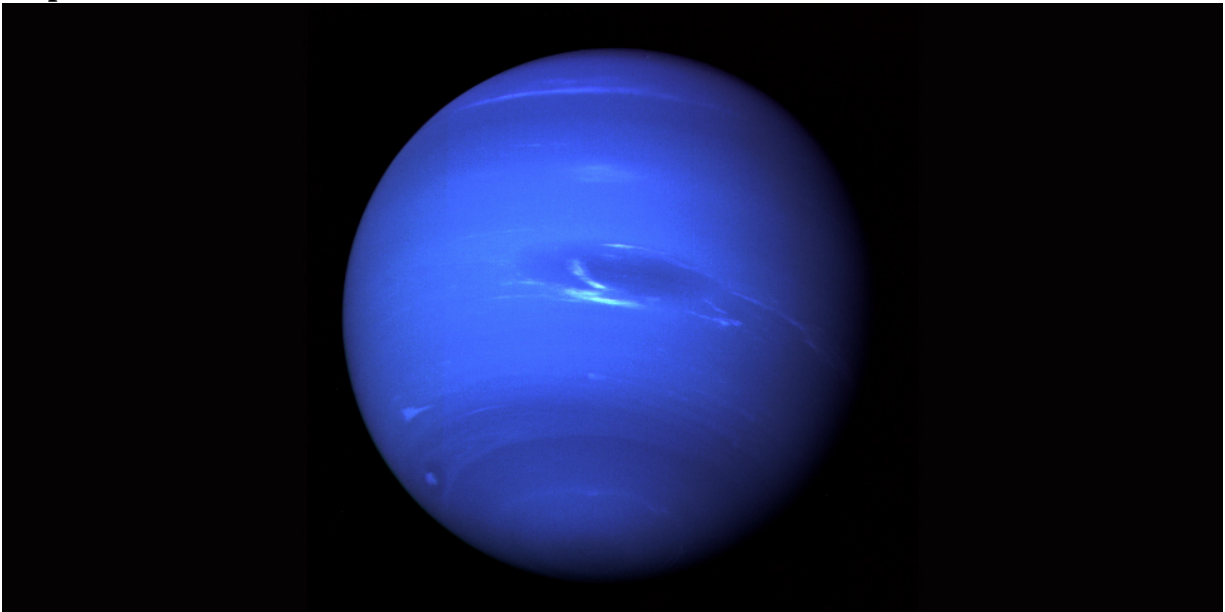
See images of [Saturn's hexagon](#) with exaggerated color in this brief NASA video.

Unlike Jupiter and Saturn, Uranus is almost entirely featureless as seen at wavelengths that range from the ultraviolet to the infrared (see its rather boring image in [\[link\]](#)). Calculations indicate that the basic atmospheric structure of Uranus should resemble that of Jupiter and Saturn, although its

upper clouds (at the 1-bar pressure level) are composed of methane rather than ammonia. However, the absence of an internal heat source suppresses up-and-down movement and leads to a very stable atmosphere with little visible structure.

Neptune differs from Uranus in its appearance, although their basic atmospheric temperatures are similar. The upper clouds are composed of methane, which forms a thin cloud layer near the top of the troposphere at a temperature of 70 K and a pressure of 1.5 bars. Most of the atmosphere above this level is clear and transparent, with less haze than is found on Uranus. The scattering of sunlight by gas molecules lends Neptune a pale blue color similar to that of Earth's atmosphere ([link](#)). Another cloud layer, perhaps composed of hydrogen sulfide ice particles, exists below the methane clouds at a pressure of 3 bars.

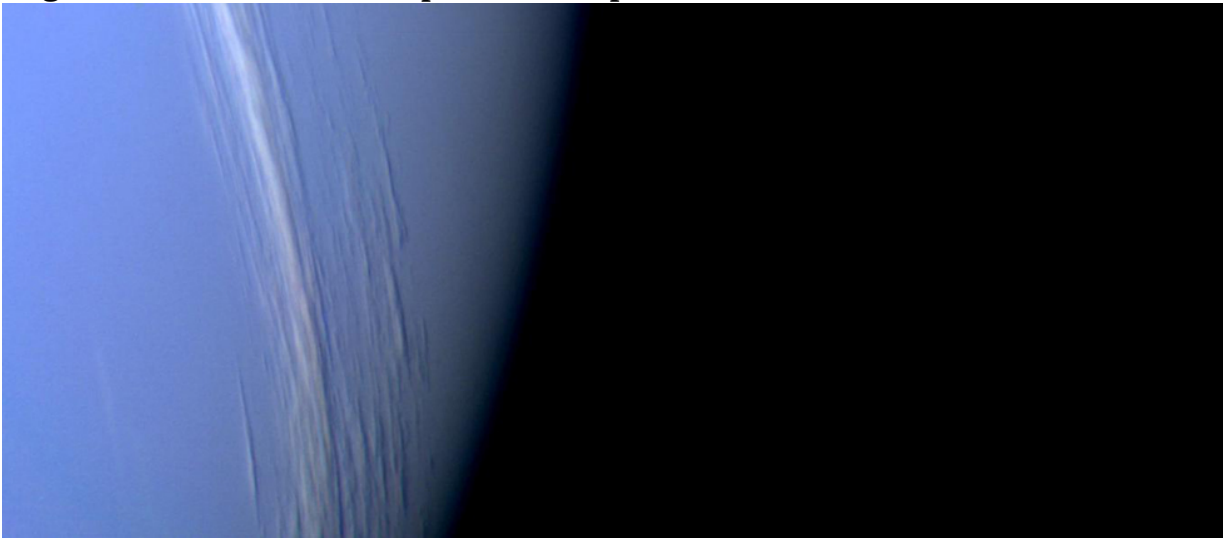
Neptune.



The planet Neptune is seen here as photographed by Voyager in 1989. The blue color, exaggerated with computer processing, is caused by the scattering of sunlight in the planet's upper atmosphere. (credit: modification of work by NASA)

Unlike Uranus, Neptune has an atmosphere in which convection currents—vertical drafts of gas—emanate from the interior, powered by the planet’s internal heat source. These currents carry warm gas above the 1.5-bar cloud level, forming additional clouds at elevations about 75 kilometers higher. These high-altitude clouds form bright white patterns against the blue planet beneath. Voyager photographed distinct shadows on the methane cloud tops, permitting the altitudes of the high clouds to be calculated. [\[link\]](#) is a remarkable close-up of Neptune’s outer layers that could never have been obtained from Earth.

High Clouds in the Atmosphere of Neptune.



These bright, narrow cirrus clouds are made of methane ice crystals. From the shadows they cast on the thicker cloud layer below, we can measure that they are about 75 kilometers higher than the main clouds.
(credit: modification of work by NASA/JPL)

Winds and Weather

The atmospheres of the jovian planets have many regions of high pressure (where there is more air) and low pressure (where there is less). Just as it does on Earth, air flows between these regions, setting up wind patterns that are then distorted by the rotation of the planet. By observing the changing

cloud patterns on the jovian planets, we can measure wind speeds and track the circulation of their atmospheres.

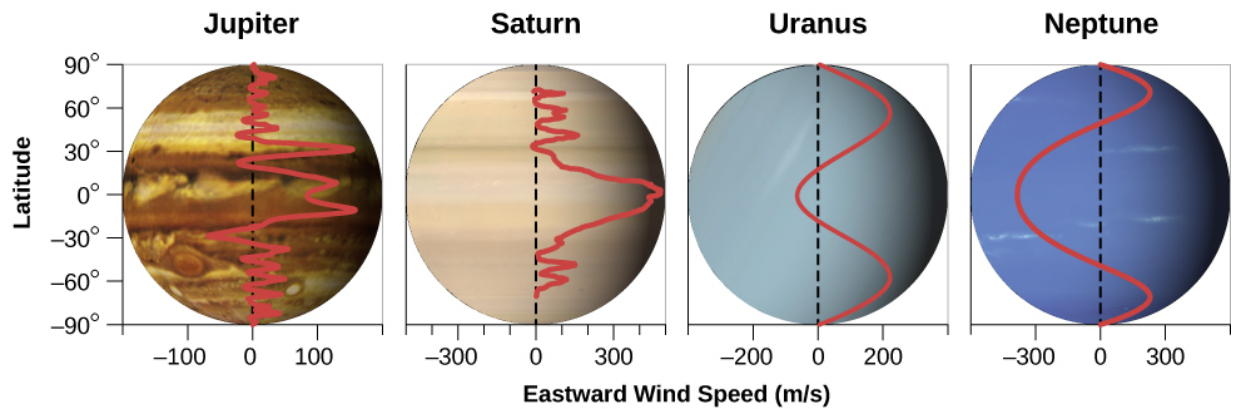
The atmospheric motions we see on these planets are fundamentally different from those on the terrestrial planets. The giants spin faster, and their rapid rotation tends to smear out of the circulation into horizontal (east-west) patterns parallel to the equator. In addition, there is no solid surface below the atmosphere against which the circulation patterns can rub and lose energy (which is how tropical storms on Earth ultimately die out when they come over land).

As we have seen, on all the giants except Uranus, heat from the inside contributes about as much energy to the atmosphere as sunlight from the outside. This means that deep convection currents of rising hot air and falling cooler air circulate throughout the atmospheres of the planets in the vertical direction.

The main features of Jupiter's visible clouds (see [\[link\]](#) and [\[link\]](#), for example) are alternating dark and light bands that stretch around the planet parallel to the equator. These bands are semi-permanent features, although they shift in intensity and position from year to year. Consistent with the small tilt of Jupiter's axis, the pattern does not change with the seasons.

More fundamental than these bands are underlying east-west wind patterns in the atmosphere, which do not appear to change at all, even over many decades. These are illustrated in [\[link\]](#), which indicates how strong the winds are at each latitude for the giant planets. At Jupiter's equator, a jet stream flows eastward with a speed of about 90 meters per second (300 kilometers per hour), similar to the speed of jet streams in Earth's upper atmosphere. At higher latitudes there are alternating east- and west-moving streams, with each hemisphere an almost perfect mirror image of the other. Saturn shows a similar pattern, but with a much stronger equatorial jet stream, as we noted earlier.

Winds on the Giant Planets.



This image compares the winds of the giant planets, illustrating that wind speed (shown on the horizontal axis) and wind direction vary with latitude (shown on the vertical axis). Winds are measured relative to a planet's internal rotation speed. A positive velocity means that the winds are blowing in the same direction as, but faster than, the planet's internal rotation. A negative velocity means that the winds are blowing more slowly than the planet's internal rotation.

The light zones on Jupiter are regions of upwelling air capped by white ammonia cirrus clouds. They apparently represent the tops of upward-moving convection currents.[\[footnote\]](#) The darker belts are regions where the cooler atmosphere moves downward, completing the convection cycle; they are darker because fewer ammonia clouds mean we can see deeper into the atmosphere, perhaps down to a region of ammonium hydrosulfide (NH_4SH) clouds. The Galileo probe sampled one of the clearest of these dry downdrafts.

Recall from earlier chapters that convection is a process in which liquids, heated from underneath, have regions where hot material rises and cooler material descends. You can see convection at work if you heat oatmeal on a stovetop or watch miso soup boil.

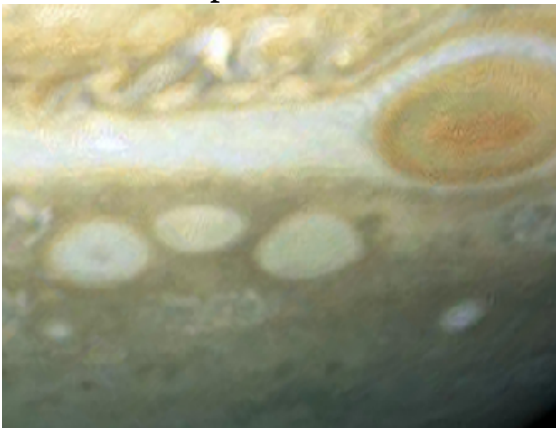
In spite of the strange seasons induced by the 98° tilt of its axis, Uranus' basic circulation is parallel with its equator, as is the case on Jupiter and Saturn. The mass of the atmosphere and its capacity to store heat are so great that the alternating 42-year periods of sunlight and darkness have little

effect. In fact, Voyager measurements show that the atmospheric temperature is even a few degrees higher on the dark winter side than on the hemisphere facing the Sun. This is another indication that the behavior of such giant planet atmospheres is a complex problem that we do not fully understand.

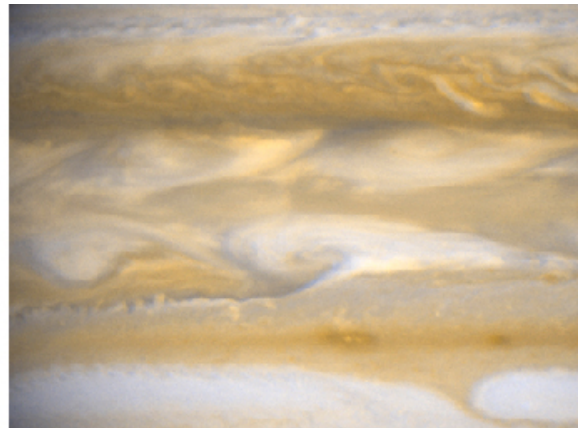
Neptune's weather is characterized by strong east-west winds generally similar to those observed on Jupiter and Saturn. The highest wind speeds near its equator reach 2100 kilometers per hour, even higher than the peak winds on Saturn. The Neptune equatorial jet stream actually approaches supersonic speeds (faster than the speed of sound in Neptune's air).

Giant Storms on Giant Planets

Superimposed on the regular atmospheric circulation patterns we have just described are many local disturbances—weather systems or *storms*, to borrow the term we use on Earth. The most prominent of these are large, oval-shaped, high-pressure regions on both Jupiter ([link](#)) and Neptune. Storms on Jupiter.



(a)



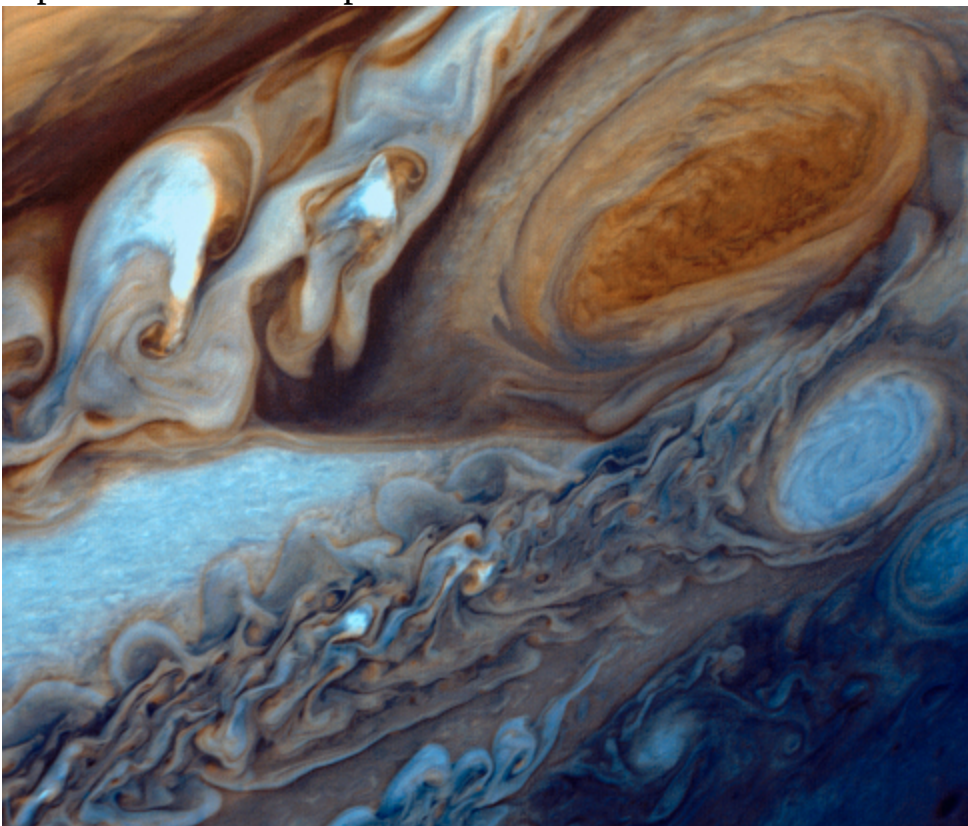
(b)

Two examples of storms on Jupiter illustrate the use of enhanced color and contrast to bring out faint features. (a) The three oval-shaped white storms below and to the left of Jupiter's Great Red Spot are highly active, and moved closer together over the course of seven months between 1994 and 1995. (b) The clouds of Jupiter are turbulent and ever-changing, as shown in this Hubble Space Telescope image from

2007. (credit a: modification of work by Reta Beebe, Amy Simon (New Mexico State Univ.), and NASA; credit b: modification of work by NASA, ESA, and A. Simon-Miller (NASA Goddard Space Flight Center))

The largest and most famous of Jupiter's storms is the Great Red Spot, a reddish oval in the southern hemisphere that changes slowly; it was 25,000 kilometers long when Voyager arrived in 1979, but it had shrunk to 20,000 kilometers by the end of the Galileo mission in 2000 ([link](#)). The giant storm has persisted in Jupiter's atmosphere ever since astronomers were first able to observe it after the invention of the telescope, more than 300 years ago. However, it has continued to shrink, raising speculation that we may see its end within a few decades.

Jupiter's Great Red Spot.



This is the largest storm system on Jupiter, as seen during the Voyager spacecraft flyby. Below and to the

right of the Red Spot is one of the white ovals, which are similar but smaller high-pressure features. The white oval is roughly the size of planet Earth, to give you a sense of the huge scale of the weather patterns we are seeing. The colors on the Jupiter image have been somewhat exaggerated here so astronomers (and astronomy students) can study their differences more effectively. See [\[link\]](#) to get a better sense of the colors your eye would actually see near Jupiter. (credit: NASA/JPL)

In addition to its longevity, the Red Spot differs from terrestrial storms in being a high-pressure region; on our planet, such storms are regions of lower pressure. The Red Spot's counterclockwise rotation has a period of six days. Three similar but smaller disturbances (about as big as Earth) formed on Jupiter in the 1930s. They look like white ovals, and one can be seen clearly below and to the right of the Great Red Spot in [\[link\]](#). In 1998, the Galileo spacecraft watched as two of these ovals collided and merged into one.

We don't know what causes the Great Red Spot or the white ovals, but we do have an idea how they can last so long once they form. On Earth, the lifetime of a large oceanic hurricane or typhoon is typically a few weeks, or even less when it moves over the continents and encounters friction with the land. Jupiter has no solid surface to slow down an atmospheric disturbance; furthermore, the sheer size of the disturbances lends them stability. We can calculate that on a planet with no solid surface, the lifetime of anything as large as the Red Spot should be measured in centuries, while lifetimes for the white ovals should be measured in decades, which is pretty much what we have observed.

Despite Neptune's smaller size and different cloud composition, Voyager showed that it had an atmospheric feature surprisingly similar to Jupiter's Great Red Spot. Neptune's Great Dark Spot was nearly 10,000 kilometers long ([\[link\]](#)). On both planets, the giant storms formed at latitude 20° S, had

the same shape, and took up about the same fraction of the planet's diameter. The Great Dark Spot rotated with a period of 17 days, versus about 6 days for the Great Red Spot. When the Hubble Space Telescope examined Neptune in the mid-1990s, however, astronomers could find no trace of the Great Dark Spot on their images.

Although many of the details of the weather on the jovian planets are not yet understood, it is clear that if you are a fan of dramatic weather, these worlds are the place to look. We study the features in these atmospheres not only for what they have to teach us about conditions in the jovian planets, but also because we hope they can help us understand the weather on Earth just a bit better.

Example:**Storms and Winds**

The wind speeds in circular storm systems can be formidable on both Earth and the giant planets. Think about our big terrestrial hurricanes. If you watch their behavior in satellite images shown on weather outlets, you will see that they require about one day to rotate. If a storm has a diameter of 400 km and rotates once in 24 h, what is the wind speed?

Solution

Speed equals distance divided by time. The distance in this case is the circumference ($2\pi R$ or πd), or approximately 1250 km, and the time is 24 h, so the speed at the edge of the storm would be about 52 km/h. Toward the center of the storm, the wind speeds can be much higher.

Check Your Learning

Jupiter's Great Red Spot rotates in 6 d and has a circumference equivalent to a circle with radius 10,000 km. Calculate the wind speed at the outer edge of the spot.

Note:**Answer:**

For the Great Red Spot of Jupiter, the circumference ($2\pi R$) is about 63,000 km. Six d equals 144 h, suggesting a speed of about 436 km/h. This is

much faster than wind speeds on Earth.

Key Concepts and Summary

The four giant planets have generally similar atmospheres, composed mostly of hydrogen and helium. Their atmospheres contain small quantities of methane and ammonia gas, both of which also condense to form clouds. Deeper (invisible) cloud layers consist of water and possibly ammonium hydrosulfide (Jupiter and Saturn) and hydrogen sulfide (Neptune). In the upper atmospheres, hydrocarbons and other trace compounds are produced by photochemistry. We do not know exactly what causes the colors in the clouds of Jupiter. Atmospheric motions on the giant planets are dominated by east-west circulation. Jupiter displays the most active cloud patterns, with Neptune second. Saturn is generally bland, in spite of its extremely high wind speeds, and Uranus is featureless (perhaps due to its lack of an internal heat source). Large storms (oval-shaped high-pressure systems such as the Great Red Spot on Jupiter and the Great Dark Spot on Neptune) can be found in some of the planet atmospheres.

For Further Exploration

Articles

Jupiter

Aguirre, Edwin. "Hubble Zooms in on Jupiter's New Red Spot." *Sky & Telescope* (August 2006): 26.

Beatty, J. "Into the Giant." *Sky & Telescope* (April 1996): 20. On the Galileo probe.

Beebe, R. "Queen of the Giant Storms." *Sky & Telescope* (October 1990): 359. Excellent review of the Red Spot.

Johnson, T. "The Galileo Mission to Jupiter and Its Moons." *Scientific American* (February 2000): 40. Results about Jupiter, Io, Ganymede, and Callisto.

Simon, A. "The Not-So-Great Red Spot." *Sky & Telescope* (March 2016): 18. On how the huge storm on Jupiter is evolving with time.

Smith, B. "Voyage of the Century." *National Geographic* (August 1990): 48. Beautiful summary of the Voyager mission to all four outer planets.

Stern, S. "Jupiter Up Close and Personal." *Astronomy* (August 2007): 28. On the New Horizons mission flyby in February 2007.

Saturn

Gore, R. "The Riddle of the Rings." *National Geographic* (July 1981): 3. Colorful report on the Voyager mission.

McEwen, A. "Cassini Unveils Saturn." *Astronomy* (July 2006): 30. A report on the first two years of discoveries in the Saturn system.

Spilker, L. "Saturn Revolution." *Astronomy* (October 2008): 34. On results from the Cassini mission.

Talcott, R. "Saturn's Sweet Surprises." *Astronomy* (June 2007): 52. On Cassini mission results.

Uranus and Neptune

Cowling, T. "Big Blue: The Twin Worlds of Uranus and Neptune." *Astronomy* (October 1990): 42. Nice, long review of the two planets.

Gore, R. "Neptune: Voyager's Last Picture Show." *National Geographic* (August 1990): 35.

Lunine, J. "Neptune at 150." *Sky & Telescope* (September 1996): 38. Nice review.

Websites

Jupiter

NASA Solar System Exploration:
<http://Solarsystem.nasa.gov/planets/jupiter>

Nine Planets Site: <http://nineplanets.org/jupiter.html>

Planetary Sciences Site:
<http://nssdc.gsfc.nasa.gov/planetary/planets/jupiterpage.html>

Saturn

NASA Solar System Exploration:
<http://Solarsystem.nasa.gov/planets/saturn>

Nine Planets Site: <http://nineplanets.org/saturn.html>

Planetary Sciences Site:
<http://nssdc.gsfc.nasa.gov/planetary/planets/saturnpage.html>

Uranus

NASA Solar System Exploration:
<http://Solarsystem.nasa.gov/planets/uranus>

Nine Planets Site: <http://nineplanets.org/uranus.html>

Planetary Sciences Site:

<http://nssdc.gsfc.nasa.gov/planetary/planets/uranuspage.html>

Neptune

NASA Solar System Exploration:

<http://Solarsystem.nasa.gov/planets/neptune>

Nine Planets Site: <http://nineplanets.org/neptune.html>

Planetary Sciences Site:

<http://nssdc.gsfc.nasa.gov/planetary/planets/neptunepage.html>

Missions

Cassini Mission Site at the Jet Propulsion Lab:

<http://saturn.jpl.nasa.gov/index.cfm>

Cassini-Huygens Mission Site at European Space Agency:

<http://sci.esa.int/cassini-huygens/>

NASA Galileo Mission Site: <http://Solarsystem.nasa.gov/galileo/>

NASA's Juno Mission to Jupiter:

http://www.nasa.gov/mission_pages/juno/main/index.html

Voyager Mission Site at the Jet Propulsion Lab: <http://voyager.jpl.nasa.gov/>

Videos

Cassini: 15 Years of Exploration: https://www.youtube.com/watch?v=2z8fzz_MBAw. Quick visual summary of mission highlights (2:29).

In the Land of Enchantment: The Epic Story of the Cassini Mission to Saturn: <https://www.youtube.com/watch?v=Vx135n8VFxY>. An inspiring illustrated lecture by Cassini Mission Imagining Lead Scientist Carolyn Porco (1:37:52).

Jupiter: The Largest Planet: <http://www.youtube.com/watch?v=s56pxa9lpvo>. Produced by NASA's Goddard Space Flight Center and Science on a Sphere (7:29).

Collaborative Group Activities

- A. A new member of Congress has asked your group to investigate why the Galileo probe launched into the Jupiter atmosphere in 1995 survived only 57 minutes and whether this was an example of a terrible scandal. Make a list of all the reasons the probe did not last longer, and why it was not made more durable. (Remember that the probe had to hitch a ride to Jupiter!)
- B. Select one of the jovian planets and organize your group to write a script for an evening news weather report for the planet you chose. Be sure you specify roughly how high in the atmosphere the region lies for which you are giving the report.
- C. What does your group think should be the next step to learn more about the giant planets? Put cost considerations aside for a moment: What kind of mission would you recommend to NASA to learn more about these giant worlds? Which world or worlds should get the highest priority and why?
- D. Suppose that an extremely dedicated (and slightly crazy) astronomer volunteers to become a human probe into Jupiter (and somehow manages to survive the trip through Jupiter's magnetosphere alive). As she enters the upper atmosphere of Jupiter, would she fall faster or slower than she would fall doing the same suicidal jump into the atmosphere of solid Earth? Groups that have some algebra background could even calculate the force she would feel compared to the force on Earth. (Bonus question: If she were in a capsule, falling into Jupiter feet first, and the floor of the capsule had a scale, what would the scale show as her weight compared to her weight on Earth?)

E. Would you or anyone in your group volunteer for a one-way, life-long mission to a space station orbiting any of the gas giants without ever being able to return to Earth? What are the challenges of such a mission? Should we leave all exploration of the outer solar system to unmanned space probes?

Review Questions

Exercise:

Problem:

What are the main challenges involved in sending probes to the giant planets?

Exercise:

Problem:

Why is it difficult to drop a probe like Galileo? How did engineers solve this problem?

Exercise:

Problem:

Explain why visual observation of the gas giants is not sufficient to determine their rotation periods, and what evidence was used to deduce the correct periods.

Exercise:

Problem: What are the seasons like on Jupiter?

Exercise:

Problem:

What is the consequence of Uranus' spin axis being 98° away from perpendicular to its orbital plane?

Exercise:

Problem: Describe the seasons on the planet Uranus.

Exercise:

Problem:

At the pressures in Jupiter's interior, describe the physical state of the hydrogen found there.

Exercise:

Problem:

Which of the gas giants has the largest icy/rocky core compared to its overall size?

Exercise:

Problem:

In the context of the giant planets and the conditions in their interiors, what is meant by "rock" and "ice"?

Exercise:

Problem: What is the primary source of Jupiter's internal heat?

Exercise:

Problem: Describe the interior heat source of Saturn.

Exercise:

Problem:

Which planet has the strongest magnetic field, and hence the largest magnetosphere? What is its source?

Exercise:

Problem:

What are the visible clouds on the four giant planets composed of, and why are they different from each other?

Exercise:

Problem:

Compare the atmospheric circulation (weather) of the four giant planets.

Exercise:

Problem:

What are the main atmospheric heat sources of each of the giant planets?

Exercise:

Problem:

Why do the upper levels of Neptune's atmosphere appear blue?

Exercise:

Problem:

How do storms on Jupiter differ from storm systems on Earth?

Thought Questions

Exercise:

Problem:

Describe the differences in the chemical makeup of the inner and outer parts of the solar system. What is the relationship between what the planets are made of and the temperature where they formed?

Exercise:

Problem: How did the giant planets grow to be so large?

Exercise:

Problem:

Jupiter is denser than water, yet composed for the most part of two light gases, hydrogen and helium. What makes Jupiter as dense as it is?

Exercise:

Problem:

Would you expect to find free oxygen gas in the atmospheres of the giant planets? Why or why not?

Exercise:

Problem:

Why would a tourist brochure (of the future) describing the most dramatic natural sights of the giant planets have to be revised more often than one for the terrestrial planets?

Exercise:

Problem:

The water clouds believed to be present on Jupiter and Saturn exist at temperatures and pressures similar to those in the clouds of the terrestrial atmosphere. What would it be like to visit such a location on Jupiter or Saturn? In what ways would the environment differ from that in the clouds of Earth?

Exercise:

Problem:

Describe the different processes that lead to substantial internal heat sources for Jupiter and Saturn. Since these two objects generate much of their energy internally, should they be called stars instead of planets? Justify your answer.

Exercise:**Problem:**

Research the Galileo mission. What technical problems occurred between the mission launch and the arrival of the craft in Jupiter's system, and how did the mission engineers deal with them? (Good sources of information include *Astronomy* and *Sky & Telescope* articles, plus the mission website.)

Figuring for Yourself**Exercise:****Problem:**

How many times more pressure exists in the interior of Jupiter compared to that of Earth?

Exercise:**Problem:**

Calculate the wind speed at the edge of Neptune's Great Dark Spot, which was 10,000 km in diameter and rotated in 17 d.

Exercise:**Problem:**

Calculate how many Earths would fit into the volumes of Saturn, Uranus, and Neptune.

Exercise:

Problem:

As the Voyager spacecraft penetrated into the outer solar system, the illumination from the Sun declined. Relative to the situation at Earth, how bright is the sunlight at each of the jovian planets?

Exercise:**Problem:**

The ions in the inner parts of Jupiter's magnetosphere rotate with the same period as Jupiter. Calculate how fast they are moving at the orbit of Jupiter's moon Io (see [Appendix G](#)). Will these ions strike Io from behind or in front as it moves about Jupiter?

Glossary

photochemistry

chemical changes caused by electromagnetic radiation

The Structure and Composition of the Sun

LEARNING OBJECTIVES

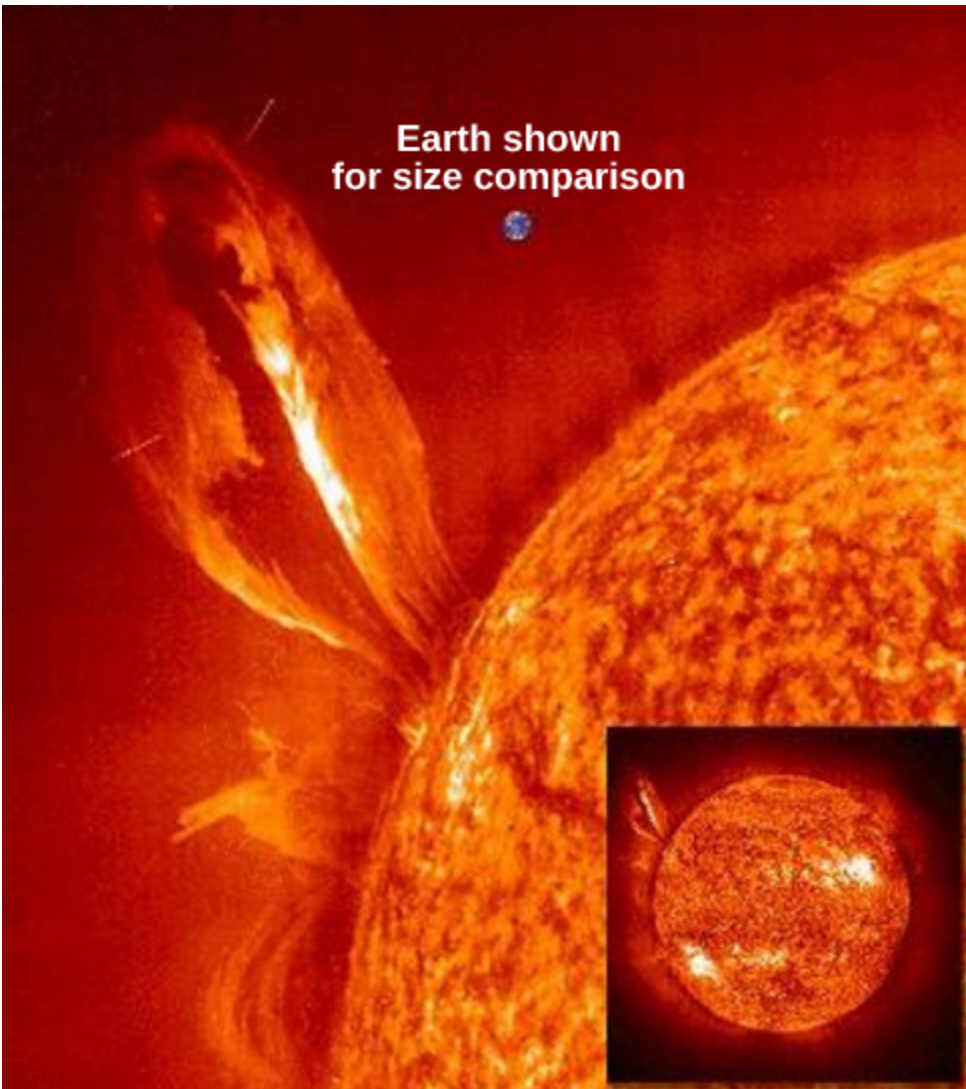
By the end of this section, you will be able to:

- Explain how the composition of the Sun differs from that of Earth
- Describe the various layers of the Sun and their functions
- Explain what happens in the different parts of the Sun's atmosphere

The Sun, like all stars, is an enormous ball of extremely hot, largely ionized gas, shining under its own power. And we do mean enormous. The Sun could fit 109 Earths side-by-side across its diameter, and it has enough volume (takes up enough space) to hold about 1.3 million Earths.

The Sun does not have a solid surface or continents like Earth, nor does it have a solid core ([\[link\]](#)). However, it does have a lot of structure and can be discussed as a series of layers, not unlike an onion. In this section, we describe the huge changes that occur in the Sun's extensive interior and atmosphere, and the dynamic and violent eruptions that occur daily in its outer layers.

Earth and the Sun.



Here, Earth is shown to scale with part of the Sun and a giant loop of hot gas erupting from its surface. The inset shows the entire Sun, smaller. (credit: modification of work by SOHO/EIT/ESA)

Some of the basic characteristics of the Sun are listed in [\[link\]](#). Although some of the terms in that table may be unfamiliar to you right now, you will get to know them as you read further.

Characteristics of the Sun		
Characteristic	How Found	Value
Mean distance	Radar reflection from planets	1 AU (149,597,892 km)
Maximum distance from Earth		1.521×10^8 km
Minimum distance from Earth		1.471×10^8 km
Mass	Orbit of Earth	333,400 Earth masses (1.99×10^{30} kg)
Mean angular diameter	Direct measure	31'59''.3
Diameter of photosphere	Angular size and distance	$109.3 \times$ Earth diameter (1.39×10^6 km)
Mean density	Mass/volume	1.41 g/cm^3 (1400 kg/m^3)
Gravitational acceleration at photosphere (surface gravity)	GM/R^2	$27.9 \times$ Earth surface gravity = 273 m/s^2

Characteristics of the Sun		
Characteristic	How Found	Value
Solar constant	Instrument sensitive to radiation at all wavelengths	1370 W/m ²
Luminosity	Solar constant × area of spherical surface 1 AU in radius	3.8×10^{26} W
Spectral class	Spectrum	G2V
Effective temperature	Derived from luminosity and radius of the Sun	5800 K
Rotation period at equator	Sunspots and Doppler shift in spectra taken at the edge of the Sun	24 days 16 hours
Inclination of equator to ecliptic	Motions of sunspots	7°10′.5

Composition of the Sun's Atmosphere

Let's begin by asking what the solar atmosphere is made of. As explained in [Radiation and Spectra](#), we can use a star's *absorption line spectrum* to determine what elements are present. It turns out that the Sun contains the same elements as Earth but *not* in the same proportions. About 73% of the Sun's mass is hydrogen, and another 25% is helium. All the other chemical elements (including those we know and love in our own bodies, such as carbon, oxygen, and nitrogen) make up only 2% of our star. The 10 most abundant gases in the Sun's visible surface layer are listed in [\[link\]](#).

Examine that table and notice that the composition of the Sun's outer layer

is very different from Earth's crust, where we live. (In our planet's crust, the three most abundant elements are oxygen, silicon, and aluminum.) Although not like our planet's, the makeup of the Sun is quite typical of stars in general.

The Abundance of Elements in the Sun		
Element	Percentage by Number of Atoms	Percentage By Mass
Hydrogen	92.0	73.4
Helium	7.8	25.0
Carbon	0.02	0.20
Nitrogen	0.008	0.09
Oxygen	0.06	0.80
Neon	0.01	0.16
Magnesium	0.003	0.06
Silicon	0.004	0.09
Sulfur	0.002	0.05
Iron	0.003	0.14

The fact that our Sun and the stars all have similar compositions and are made up of mostly hydrogen and helium was first shown in a brilliant thesis

in 1925 by Cecilia Payne-Gaposchkin, the first woman to get a PhD in astronomy in the United States ([\[link\]](#)). However, the idea that the simplest light gases—hydrogen and helium—were the most abundant elements in stars was so unexpected and so shocking that she assumed her analysis of the data must be wrong. At the time, she wrote, “The enormous abundance derived for these elements in the stellar atmosphere is almost certainly not real.” Even scientists sometimes find it hard to accept new ideas that do not agree with what everyone “knows” to be right.

Cecilia Payne-Gaposchkin (1900–1979).



Her 1925 doctoral thesis laid the foundations for understanding the composition of the Sun and the stars. Yet, being a woman, she was not given a formal appointment at Harvard, where she worked, until 1938 and was not appointed a professor until 1956. (credit: Smithsonian Institution)

Before Payne-Gaposchkin’s work, everyone assumed that the composition of the Sun and stars would be much like that of Earth. It was 3 years after her thesis that other studies proved beyond a doubt that the enormous abundance of hydrogen and helium in the Sun is indeed real. (And, as we

will see, the composition of the Sun and the stars is much more typical of the makeup of the universe than the odd concentration of heavier elements that characterizes our planet.)

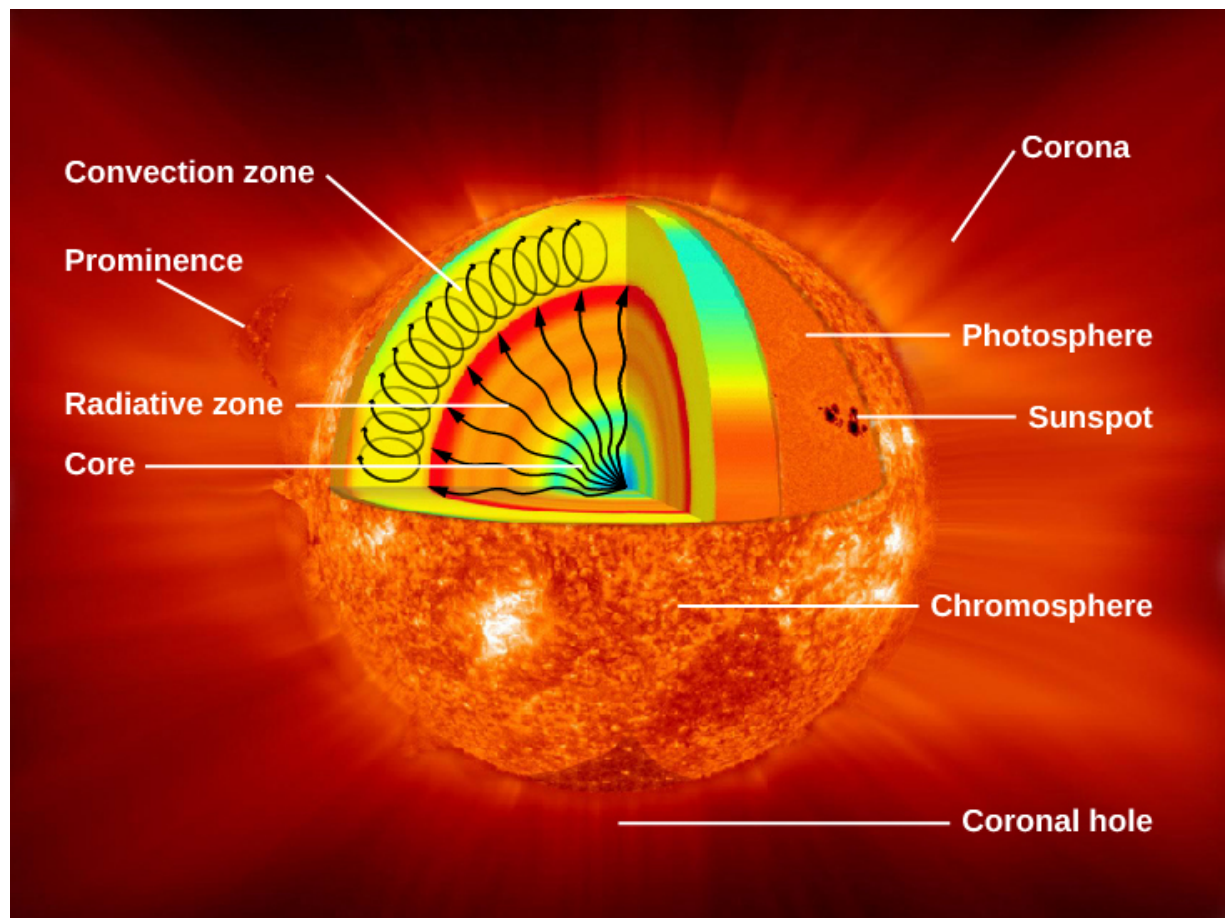
Most of the elements found in the Sun are in the form of atoms, with a small number of molecules, all in the form of gases: the Sun is so hot that no matter can survive as a liquid or a solid. In fact, the Sun is so hot that many of the atoms in it are *ionized*, that is, stripped of one or more of their electrons. This removal of electrons from their atoms means that there is a large quantity of free electrons and positively charged ions in the Sun, making it an electrically charged environment—quite different from the neutral one in which you are reading this text. (Scientists call such a hot ionized gas a **plasma**.)

In the nineteenth century, scientists observed a spectral line at 530.3 nanometers in the Sun's outer atmosphere, called the corona (a layer we will discuss in a minute.) This line had never been seen before, and so it was assumed that this line was the result of a new element found in the corona, quickly named coronium. It was not until 60 years later that astronomers discovered that this emission was in fact due to highly ionized iron—iron with 13 of its electrons stripped off. This is how we first discovered that the Sun's atmosphere had a temperature of more than a million degrees.

The Layers of the Sun beneath the Visible Surface

[\[link\]](#) shows what the Sun would look like if we could see all parts of it from the center to its outer atmosphere; the terms in the figure will become familiar to you as you read on.

Parts of the Sun.



This illustration shows the different parts of the Sun, from the hot core where the energy is generated through regions where energy is transported outward, first by radiation, then by convection, and then out through the solar atmosphere. The parts of the atmosphere are also labeled the photosphere, chromosphere, and corona. Some typical features in the atmosphere are shown, such as coronal holes and prominences. (credit: modification of work by NASA/Goddard)

The Sun's layers are different from each other, and each plays a part in producing the energy that the Sun ultimately emits. We will begin with the core and work our way out through the layers. The Sun's *core* is extremely dense and is the source of all of its energy. Inside the core, nuclear energy is being released (in ways we will discuss in [The Sun: A Nuclear Powerhouse](#)). The core is approximately 20% of the size of the solar

interior and is thought to have a temperature of approximately 15 million K, making it the hottest part of the Sun.

Above the core is a region known as the *radiative zone*—named for the primary mode of transporting energy across it. This region starts at about 25% of the distance to the solar surface and extends up to about 70% of the way to the surface. The light generated in the core is transported through the radiative zone very slowly, since the high density of matter in this region means a photon cannot travel too far without encountering a particle, causing it to change direction and lose some energy.

The *convective zone* is the outermost layer of the solar interior. It is a thick layer approximately 200,000 kilometers deep that transports energy from the edge of the radiative zone to the surface through giant convection cells, similar to a pot of boiling oatmeal. The plasma at the bottom of the convective zone is extremely hot, and it bubbles to the surface where it loses its heat to space. Once the plasma cools, it sinks back to the bottom of the convective zone.

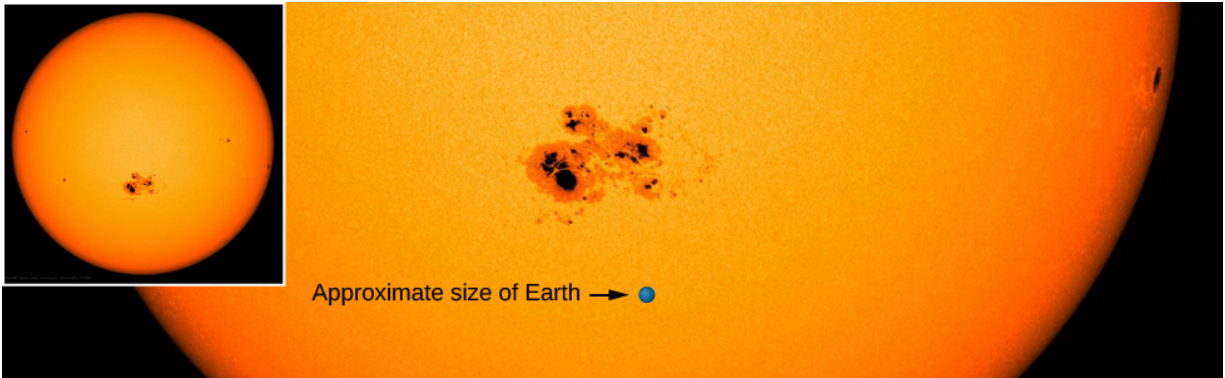
Now that we have given a quick overview of the structure of the whole Sun, in this section, we will embark on a journey through the visible layers of the Sun, beginning with the photosphere—the visible surface.

The Solar Photosphere

Earth's air is generally transparent. But on a smoggy day in many cities, it can become opaque, which prevents us from seeing through it past a certain point. Something similar happens in the Sun. Its outer atmosphere is transparent, allowing us to look a short distance through it. But when we try to look through the atmosphere deeper into the Sun, our view is blocked.

The **photosphere** is the layer where the Sun becomes opaque and marks the boundary past which we cannot see ([link](#)).

Solar Photosphere plus Sunspots.



This photograph shows the photosphere—the visible surface of the Sun. Also shown is an enlarged image of a group of sunspots; the size of Earth is shown for comparison. Sunspots appear darker because they are cooler than their surroundings. The typical temperature at the center of a large sunspot is about 3800 K, whereas the photosphere has a temperature of about 5800 K. (credit: modification of work by NASA/SDO)

As we saw, the energy that emerges from the photosphere was originally generated deep inside the Sun (more on this in [The Sun: A Nuclear Powerhouse](#)). This energy is in the form of photons, which make their way slowly toward the solar surface. Outside the Sun, we can observe *only* those photons that are emitted into the solar photosphere, where the density of atoms is sufficiently low and the photons can finally escape from the Sun without colliding with another atom or ion.

As an analogy, imagine that you are attending a big campus rally and have found a prime spot near the center of the action. Your friend arrives late and calls you on your cell phone to ask you to join her at the edge of the crowd. You decide that friendship is worth more than a prime spot, and so you work your way out through the dense crowd to meet her. You can move only a short distance before bumping into someone, changing direction, and trying again, making your way slowly to the outside edge of the crowd. All this while, your efforts are not visible to your waiting friend at the edge. Your friend can't see you until you get very close to the edge because of all the bodies in the way. So too photons making their way through the Sun are

constantly bumping into atoms, changing direction, working their way slowly outward, and becoming visible only when they reach the atmosphere of the Sun where the density of atoms is too low to block their outward progress.

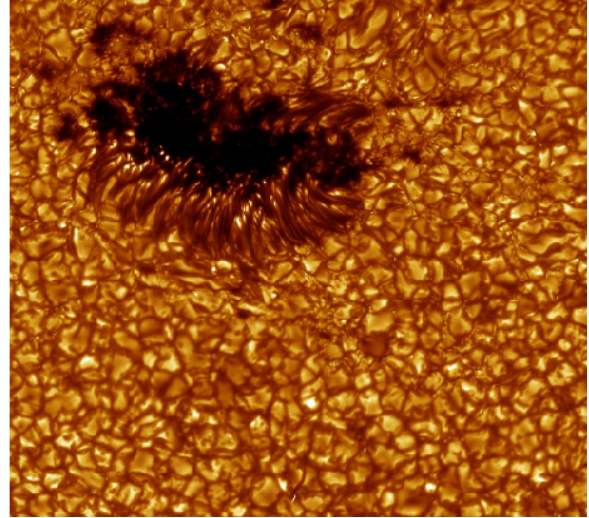
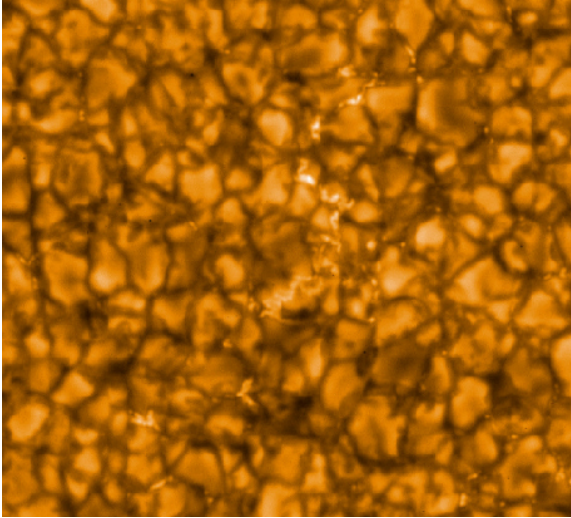
Astronomers have found that the solar atmosphere changes from almost perfectly transparent to almost completely opaque in a distance of just over 400 kilometers; it is this thin region that we call the *photosphere*, a word that comes from the Greek for “light sphere.” When astronomers speak of the “diameter” of the Sun, they mean the size of the region surrounded by the photosphere.

The photosphere looks sharp only from a distance. If you were falling into the Sun, you would not feel any surface but would just sense a gradual increase in the density of the gas surrounding you. It is much the same as falling through a cloud while skydiving. From far away, the cloud looks as if it has a sharp surface, but you do not feel a surface as you fall into it. (One big difference between these two scenarios, however, is temperature. The Sun is so hot that you would be vaporized long before you reached the photosphere. Skydiving in Earth’s atmosphere is much safer.)

We might note that the atmosphere of the Sun is not a very dense layer compared to the air in the room where you are reading this text. At a typical point in the photosphere, the pressure is less than 10% of Earth’s pressure at sea level, and the density is about one ten-thousandth of Earth’s atmospheric density at sea level.

Observations with telescopes show that the photosphere has a mottled appearance, resembling grains of rice spilled on a dark tablecloth or a pot of boiling oatmeal. This structure of the photosphere is called **granulation** (see [\[link\]](#)). Granules, which are typically 700 to 1000 kilometers in diameter (about the width of Texas), appear as bright areas surrounded by narrow, darker (cooler) regions. The lifetime of an individual granule is only 5 to 10 minutes. Even larger are supergranules, which are about 35,000 kilometers across (about the size of two Earths) and last about 24 hours.

Granulation Pattern.



The surface markings of the convection cells create a granulation pattern on this dramatic image (left) taken from the Japanese Hinode spacecraft. You can see the same pattern when you heat up miso soup. The right image shows an irregular-shaped sunspot and granules on the Sun's surface, seen with the Swedish Solar Telescope on August 22, 2003. (credit left: modification of work by Hinode JAXA/NASA/PPARC; credit right: ISP/SST/Oddbjorn Engvold, Jun Elin Wiik, Luc Rouppe van der Voort)

The motions of the granules can be studied by examining the Doppler shifts in the spectra of gases just above them (see [The Doppler Effect](#)). The bright granules are columns of hotter gases rising at speeds of 2 to 3 kilometers per second from below the photosphere. As this rising gas reaches the photosphere, it spreads out, cools, and sinks down again into the darker regions between the granules. Measurements show that the centers of the granules are hotter than the intergranular regions by 50 to 100 K.

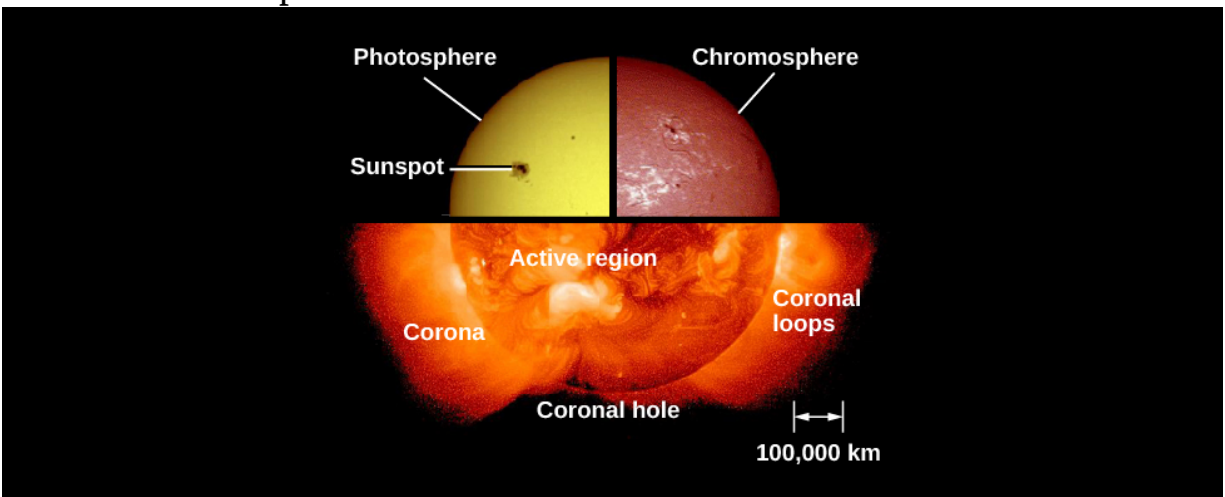
Note:

See the “boiling” action of granulation in this [30-second time-lapse video](#) from the Swedish Institute for Solar Physics.

The Chromosphere

The Sun's outer gases extend far beyond the photosphere ([\[link\]](#)). Because they are transparent to most visible radiation and emit only a small amount of light, these outer layers are difficult to observe. The region of the Sun's atmosphere that lies immediately above the photosphere is called the **chromosphere**. Until this century, the chromosphere was visible only when the photosphere was concealed by the Moon during a total solar eclipse (see the chapter on [Earth, Moon, and Sky](#)). In the seventeenth century, several observers described what appeared to them as a narrow red "streak" or "fringe" around the edge of the Moon during a brief instant after the Sun's photosphere had been covered. The name *chromosphere*, from the Greek for "colored sphere," was given to this red streak.

The Sun's Atmosphere.



Composite image showing the three components of the solar atmosphere: the photosphere or surface of the Sun taken in ordinary light; the chromosphere, imaged in the light of the strong red spectral line of hydrogen (H-alpha); and the corona as seen with X-rays.
(credit: modification of work by NASA)

Observations made during eclipses show that the chromosphere is about 2000 to 3000 kilometers thick, and its spectrum consists of bright emission lines, indicating that this layer is composed of hot gases emitting light at

discrete wavelengths. The reddish color of the chromosphere arises from one of the strongest emission lines in the visible part of its spectrum—the bright red line caused by hydrogen, the element that, as we have already seen, dominates the composition of the Sun.

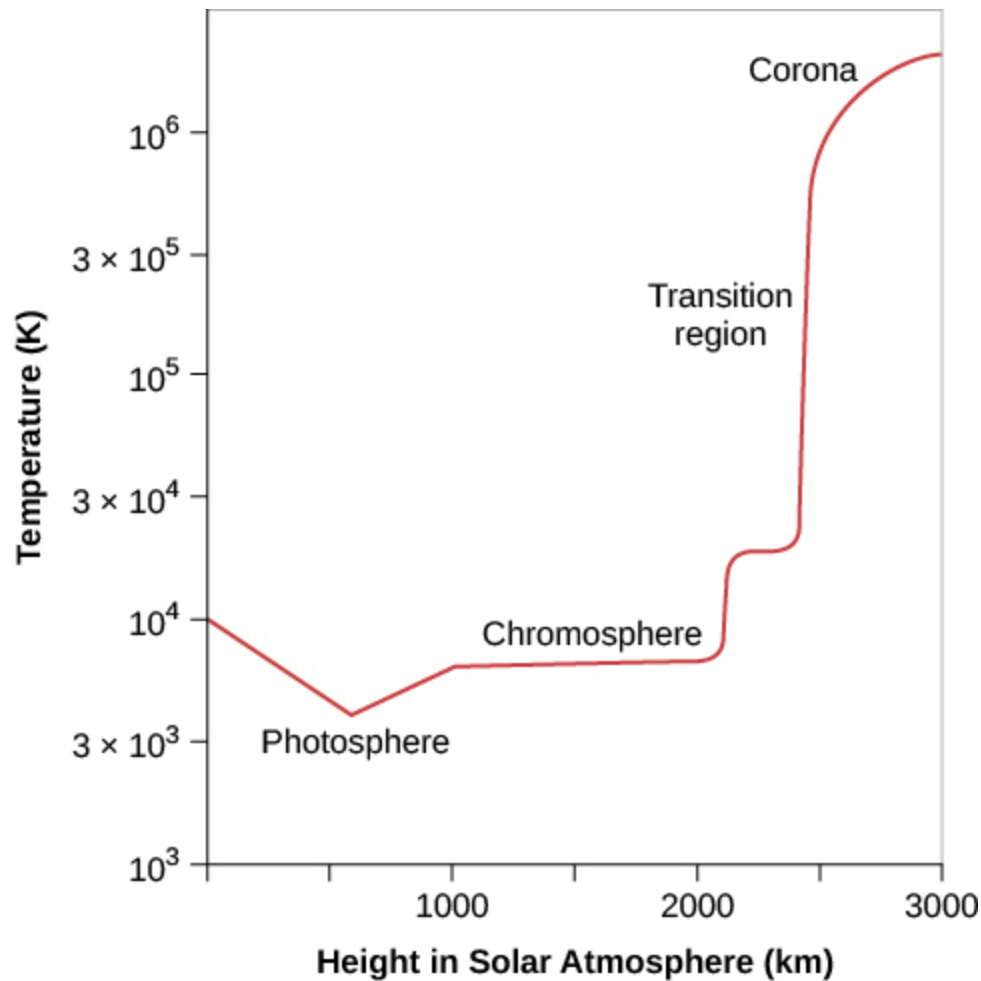
In 1868, observations of the chromospheric spectrum revealed a yellow emission line that did not correspond to any previously known element on Earth. Scientists quickly realized they had found a new element and named it *helium* (after *helios*, the Greek word for “Sun”). It took until 1895 for helium to be discovered on our planet. Today, students are probably most familiar with it as the light gas used to inflate balloons, although it turns out to be the second-most abundant element in the universe.

The temperature of the chromosphere is about 10,000 K. This means that the chromosphere is hotter than the photosphere, which should seem surprising. In all the situations we are familiar with, temperatures fall as one moves away from the source of heat, and the chromosphere is farther from the center of the Sun than the photosphere is.

The Transition Region

The increase in temperature does not stop with the chromosphere. Above it is a region in the solar atmosphere where the temperature changes from 10,000 K (typical of the chromosphere) to nearly a million degrees. The hottest part of the solar atmosphere, which has a temperature of a million degrees or more, is called the **corona**. Appropriately, the part of the Sun where the rapid temperature rise occurs is called the **transition region**. It is probably only a few tens of kilometers thick. [\[link\]](#) summarizes how the temperature of the solar atmosphere changes from the photosphere outward.

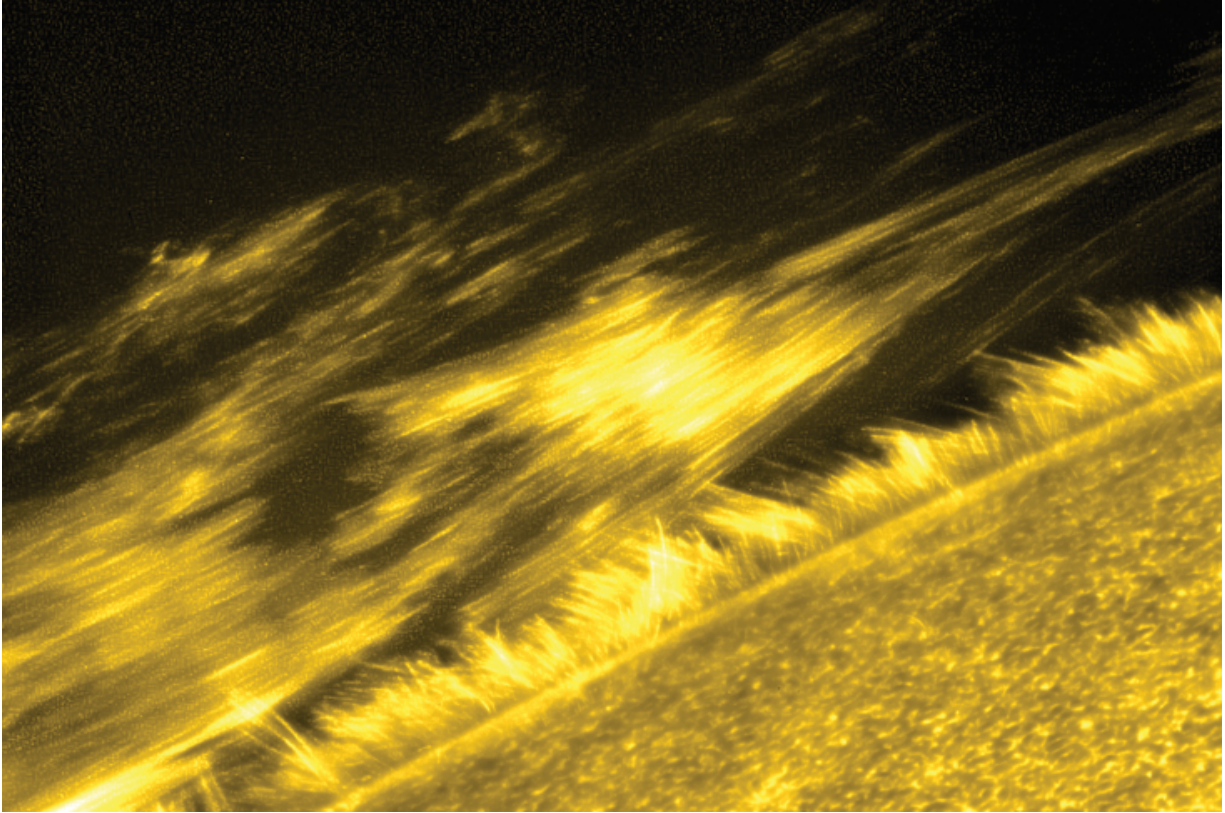
Temperatures in the Solar Atmosphere.



On this graph, temperature is shown increasing upward, and height above the photosphere is shown increasing to the right. Note the very rapid increase in temperature over a very short distance in the transition region between the chromosphere and the corona.

In 2013, NASA launched the Interface Region Imaging Spectrograph (IRIS) to study the transition region to understand better how and why this sharp temperature increase occurs. IRIS is the first space mission that is able to obtain high spatial resolution images of the different features produced over this wide temperature range and to see how they change with time and location ([link](#)).

Portion of the Transition Region.



This image shows a giant ribbon of relatively cool gas threading through the lower portion of the hot corona. This ribbon (the technical term is filament) is made up of many individual threads. Time-lapse movies of this filament showed that it gradually heated as it moved through the corona. Scientists study events like this in order to try to understand what heats the chromosphere and corona to high temperatures. The “whiskers” at the edge of the Sun are spicules, jets of gas that shoot material up from the Sun’s surface and disappear after only a few minutes. This single image gives a hint of just how complicated it is to construct a model of the all the different structures and heating mechanisms in the solar atmosphere. (credit: JAXA/NASA/Hinode)

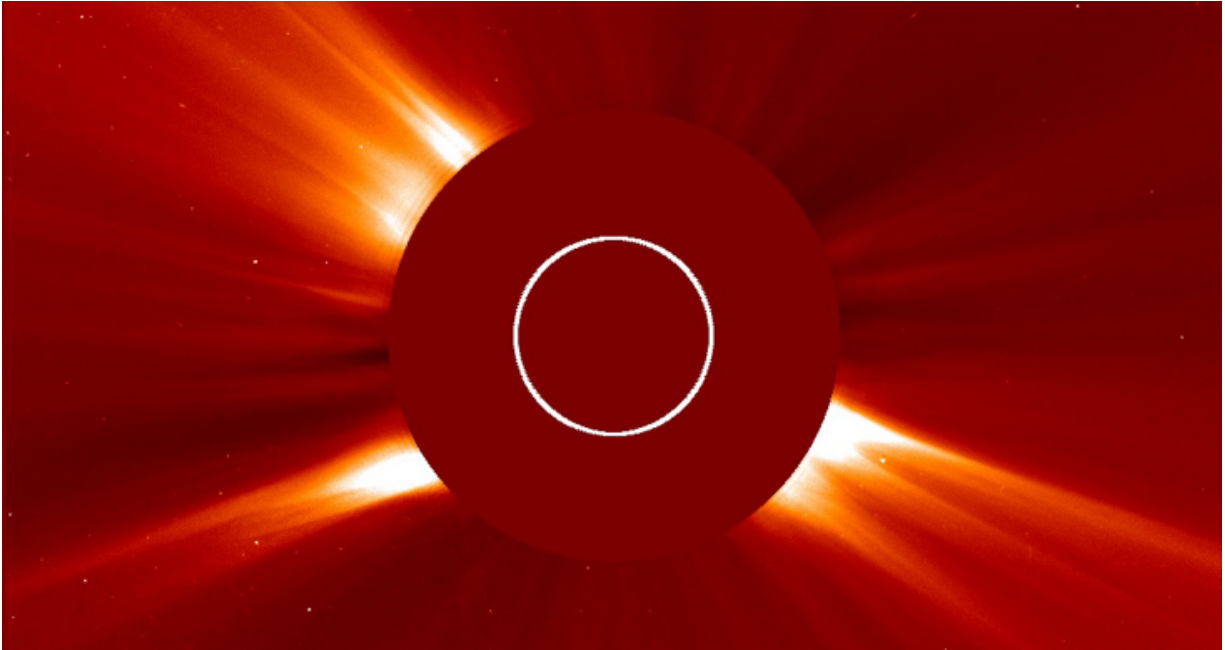
[\[link\]](#) and the red graph in [\[link\]](#) make the Sun seem rather like an onion, with smooth spherical shells, each one with a different temperature. For a long time, astronomers did indeed think of the Sun this way. However, we

now know that while this idea of layers—photosphere, chromosphere, transition region, corona—describes the big picture fairly well, the Sun's atmosphere is really more complicated, with hot and cool regions intermixed. For example, clouds of carbon monoxide gas with temperatures colder than 4000 K have now been found at the same height above the photosphere as the much hotter gas of the chromosphere.

The Corona

The outermost part of the Sun's atmosphere is called the *corona*. Like the chromosphere, the corona was first observed during total eclipses ([\[link\]](#)). Unlike the chromosphere, the corona has been known for many centuries: it was referred to by the Roman historian Plutarch and was discussed in some detail by Kepler.

The corona extends millions of kilometers above the photosphere and emits about half as much light as the full moon. The reason we don't see this light until an eclipse occurs is the overpowering brilliance of the photosphere. Just as bright city lights make it difficult to see faint starlight, so too does the intense light from the photosphere hide the faint light from the corona. While the best time to see the corona from Earth is during a total solar eclipse, it can be observed easily from orbiting spacecraft. Its brighter parts can now be photographed with a special instrument—a coronagraph—that removes the Sun's glare from the image with an occulting disk (a circular piece of material held so it is just in front of the Sun).
Coronagraph.



This image of the Sun was taken March 2, 2016. The larger dark circle in the center is the disk that blocks the Sun's glare, allowing us to see the corona. The smaller inner circle is where the Sun would be if it were visible in this image. (credit: modification of work by NASA/SOHO)

Studies of its spectrum show the corona to be very low in density. At the bottom of the corona, there are only about 10^9 atoms per cubic centimeter, compared with about 10^{16} atoms per cubic centimeter in the upper photosphere and 10^{19} molecules per cubic centimeter at sea level in Earth's atmosphere. The corona thins out very rapidly at greater heights, where it corresponds to a high vacuum by Earth laboratory standards. The corona extends so far into space—far past Earth—that here on our planet, we are technically living in the Sun's atmosphere.

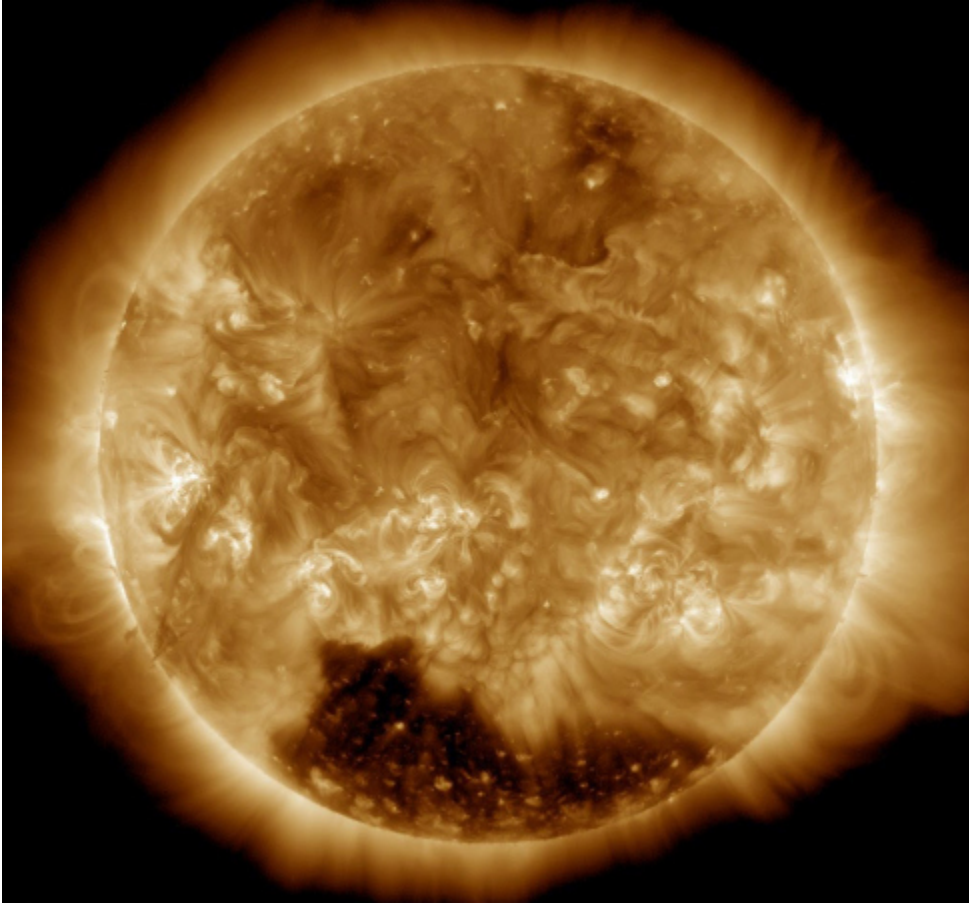
The Solar Wind

One of the most remarkable discoveries about the Sun's atmosphere is that it produces a stream of charged particles (mainly protons and electrons) that we call the **solar wind**. These particles flow outward from the Sun into the

solar system at a speed of about 400 kilometers per second (almost 1 million miles per hour)! The solar wind exists because the gases in the corona are so hot and moving so rapidly that they cannot be held back by solar gravity. (This wind was actually discovered by its effects on the charged tails of comets; in a sense, we can see the comet tails blow in the solar breeze the way wind socks at an airport or curtains in an open window flutter on Earth.)

Although the solar wind material is very, very rarified (i.e., *extremely* low density), the Sun has an enormous surface area. Astronomers estimate that the Sun is losing about 1–2 million tons of material each second through this wind. Although this sounds like a lot, it's so trivial compared to the enormous mass of the Sun that it can be neglected as we study the Sun.

From where in the Sun does the solar wind emerge? In visible photographs, the solar corona appears fairly uniform and smooth. X-ray and extreme ultraviolet pictures, however, show that the corona has loops, plumes, and both bright and dark regions. Large dark regions of the corona that are relatively cool and quiet are called **coronal holes** ([\[link\]](#)). In these regions, magnetic field lines stretch far out into space away from the Sun, rather than looping back to the surface. The solar wind comes predominantly from coronal holes, where gas can stream away from the Sun into space unhindered by magnetic fields. Hot coronal gas, on the other hand, is present mainly where magnetic fields have trapped and concentrated it.



Coronal Hole. The dark area visible near the Sun's south pole on this Solar Dynamics Observer spacecraft image is a coronal hole. (credit: modification of work by NASA/SDO)

At the surface of Earth, we are protected to some degree from the solar wind by our atmosphere and Earth's magnetic field (see [Earth as a Planet](#)). However, the magnetic field lines come into Earth at the north and south magnetic poles. Here, charged particles accelerated by the solar wind can follow the field down into our atmosphere. As the particles strike molecules of air, they cause them to glow, producing beautiful curtains of light called the **auroras**, or the northern and southern lights ([link](#)).
Aurora.



The colorful glow in the sky results from charged particles in a solar wind interacting with Earth's magnetic fields. The stunning display captured here occurred over Jokulsarlon Lake in Iceland in 2013.
(credit: Moyan Brenn)

Note:

This [NASA video](#) explains and demonstrates the nature of the auroras and their relationship to Earth's magnetic field.

Key Concepts and Summary

The Sun, our star, has several layers beneath the visible surface: the core, radiative zone, and convective zone. These, in turn, are surrounded by a number of layers that make up the solar atmosphere. In order of increasing

distance from the center of the Sun, they are the photosphere, with a temperature that ranges from 4500 K to about 6800 K; the chromosphere, with a typical temperature of 10^4 K; the transition region, a zone that may be only a few kilometers thick, where the temperature increases rapidly from 10^4 K to 10^6 K; and the corona, with temperatures of a few million K. The Sun's surface is mottled with upwelling convection currents seen as hot, bright granules. Solar wind particles stream out into the solar system through coronal holes. When such particles reach the vicinity of Earth, they produce auroras, which are strongest near Earth's magnetic poles. Hydrogen and helium together make up 98% of the mass of the Sun, whose composition is much more characteristic of the universe at large than is the composition of Earth.

Glossary

aurora

light radiated by atoms and ions in the ionosphere excited by charged particles from the Sun, mostly seen in the magnetic polar regions

chromosphere

the part of the solar atmosphere that lies immediately above the photospheric layers

corona

(of the Sun) the outer (hot) atmosphere of the Sun

coronal hole

a region in the Sun's outer atmosphere that appears darker because there is less hot gas there

granulation

the rice-grain-like structure of the solar photosphere; granulation is produced by upwelling currents of gas that are slightly hotter, and therefore brighter, than the surrounding regions, which are flowing downward into the Sun

photosphere

the region of the solar (or stellar) atmosphere from which continuous radiation escapes into space

plasma

a hot ionized gas

solar wind

a flow of hot, charged particles leaving the Sun

transition region

the region in the Sun's atmosphere where the temperature rises very rapidly from the relatively low temperatures that characterize the chromosphere to the high temperatures of the corona

The Solar Cycle

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the sunspot cycle and, more generally, the solar cycle
- Explain how magnetism is the source of solar activity

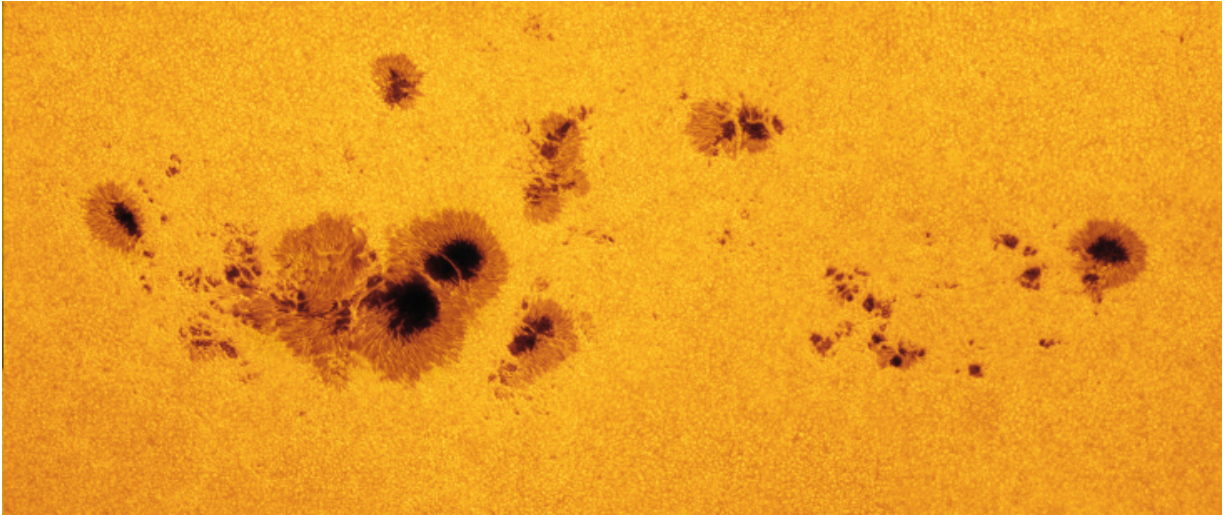
Before the invention of the telescope, the Sun was thought to be an unchanging and perfect sphere. We now know that the Sun is in a perpetual state of change: its surface is a seething, bubbling cauldron of hot gas. Areas that are darker and cooler than the rest of the surface come and go. Vast plumes of gas erupt into the chromosphere and corona. Occasionally, there are even giant explosions on the Sun that send enormous streamers of charged particles and energy hurtling toward Earth. When they arrive, these can cause power outages and other serious effects on our planet.

Sunspots

The first evidence that the Sun changes came from studies of **sunspots**, which are large, dark features seen on the surface of the Sun caused by increased magnetic activity. They look darker because the spots are typically at a temperature of about 3800 K, whereas the bright regions that surround them are at about 5800 K ([\[link\]](#)). Occasionally, these spots are large enough to be visible to the unaided eye, and we have records going back over a thousand years from observers who noticed them when haze or mist reduced the Sun's intensity. (We emphasize what your parents have surely told you: looking at the Sun for even a brief time can cause

permanent eye damage. This is the one area of astronomy where we don't encourage you to do your own observing without getting careful instructions or filters from your instructor.)

Sunspots.



This image of sunspots, cooler and thus darker regions on the Sun, was taken in July 2012. You can see the dark, central region of each sunspot (called the umbra) surrounded by a less dark region (the penumbra). The largest spot shown here is about 11 Earths wide. Although sunspots appear dark when seen next to the hotter gases of the photosphere, an average sunspot, cut out of the solar surface and left standing in the night sky, would be about as bright as the full moon. The mottled appearance of the Sun's surface is granulation.
(credit: NASA Goddard Space Flight Center, Alan Friedman)

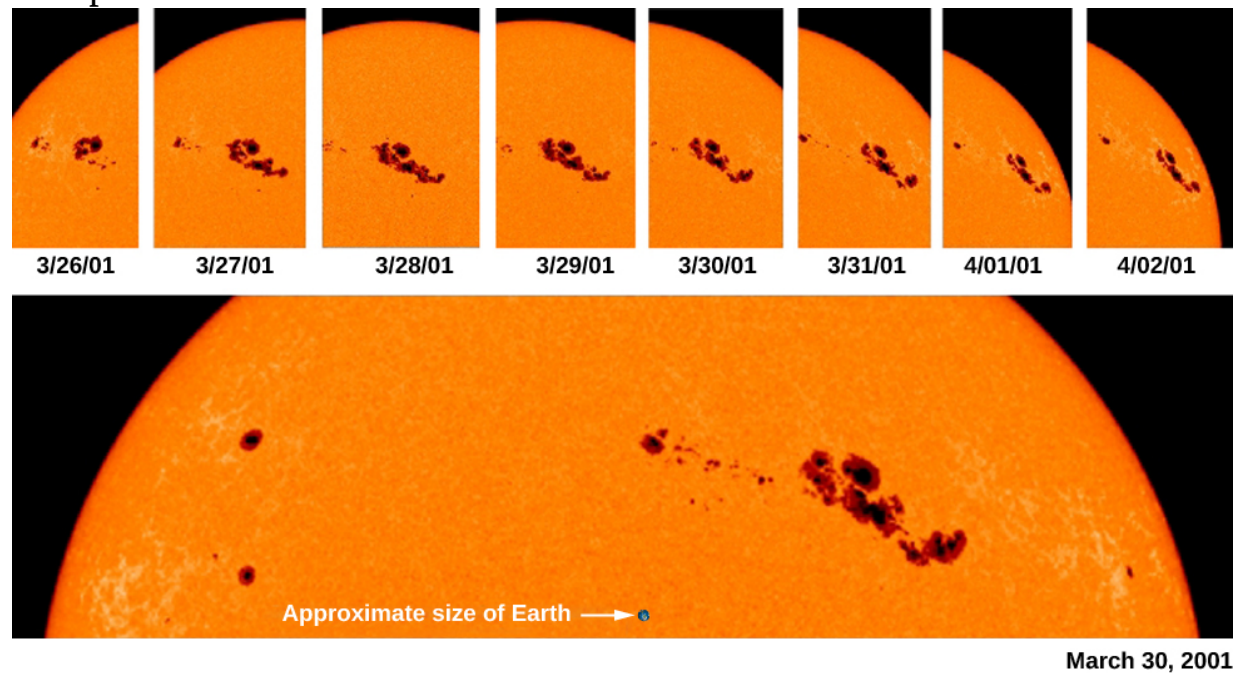
While we understand that sunspots look darker because they are cooler, they are nevertheless hotter than the surfaces of many stars. If they could be removed from the Sun, they would shine brightly. They appear dark only in contrast with the hotter, brighter photosphere around them.

Individual sunspots come and go, with lifetimes that range from a few hours to a few months. If a spot lasts and develops, it usually consists of two parts: an inner darker core, the *umbra*, and a surrounding less dark region, the *penumbra*. Many spots become much larger than Earth, and a few, like

the largest one shown in [\[link\]](#), have reached diameters over 140,000 kilometers. Frequently, spots occur in groups of 2 to 20 or more. The largest groups are very complex and may have over 100 spots. Like storms on Earth, sunspots are not fixed in position, but they drift slowly compared with the Sun's rotation.

By recording the apparent motions of the sunspots as the turning Sun carried them across its disk ([\[link\]](#)), Galileo, in 1612, demonstrated that the Sun rotates on its axis with a rotation period of approximately 1 month. Our star turns in a west-to-east direction, like the orbital motions of the planets. The Sun, however, is a gas and does not have to rotate rigidly, the way a solid body like Earth does. Modern observations show that the speed of rotation of the Sun varies according to latitude, that is, it's different as you go north or south of the Sun's equator. The rotation period is about 25 days at the equator, 28 days at latitude 40° , and 36 days at latitude 80° . We call this behavior **differential rotation**.

Sunspots Rotate Across Sun's Surface.



This sequence of photographs of the Sun's surface tracks the movement of sunspots across the visible hemisphere of the Sun. On March 30, 2001, this group of sunspots extended across an area about 13 times the diameter of Earth. This region produced many flares and

coronal mass ejections. (credit: modification of work by SOHO/NASA/ESA)

The Sunspot Cycle

Between 1826 and 1850, Heinrich Schwabe, a German pharmacist and amateur astronomer, kept daily records of the number of sunspots. What he was really looking for was a planet inside the orbit of Mercury, which he hoped to find by observing its dark silhouette as it passed between the Sun and Earth. He failed to find the hoped-for planet, but his diligence paid off with an even-more important discovery: the **sunspot cycle**. He found that the number of sunspots varied systematically, in cycles about a decade long.

What Schwabe observed was that, although individual spots are short lived, the total number visible on the Sun at any one time was likely to be very much greater at certain times—the periods of *sunspot maximum*—than at other times—the periods of *sunspot minimum*. We now know that sunspot maxima occur at an *average* interval of 11 years, but the intervals between successive maxima have ranged from as short as 9 years to as long as 14 years. During sunspot maxima, more than 100 spots can often be seen at once. Even then, less than one-half of one percent of the Sun’s surface is covered by spots ([\[link\]](#)). During sunspot minima, sometimes no spots are visible. The Sun’s activity reached its most recent maximum in 2014.

Note:

Watch this brief [video](#) from NASA’s Goddard Space Flight Center that explains the sunspot cycle.

Magnetism and the Solar Cycle

Now that we have discussed the Sun’s activity cycle, you might be asking, “Why does the Sun change in such a regular way?” Astronomers now

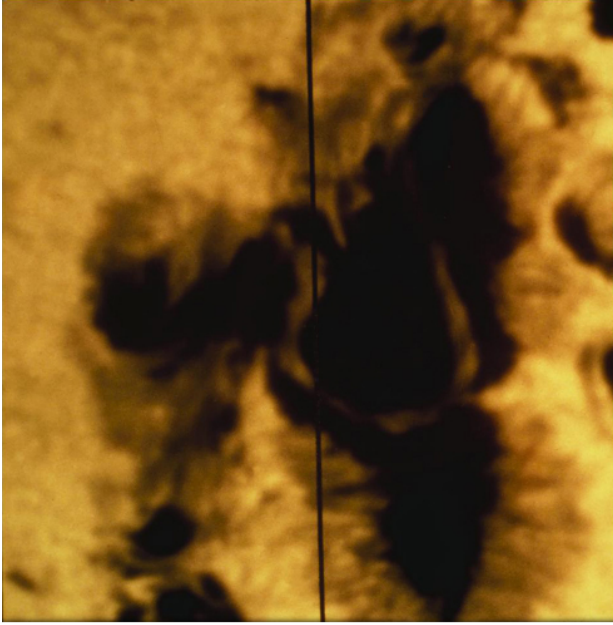
understand that it is the Sun's changing magnetic field that drives solar activity.

The solar magnetic field is measured using a property of atoms called the *Zeeman effect*. Recall from [Radiation and Spectra](#) that an atom has many energy levels and that spectral lines are formed when electrons shift from one level to another. If each energy level is precisely defined, then the difference between them is also quite precise. As an electron changes levels, the result is a sharp, narrow spectral line (either an absorption or emission line, depending on whether the electron's energy increases or decreases in the transition).

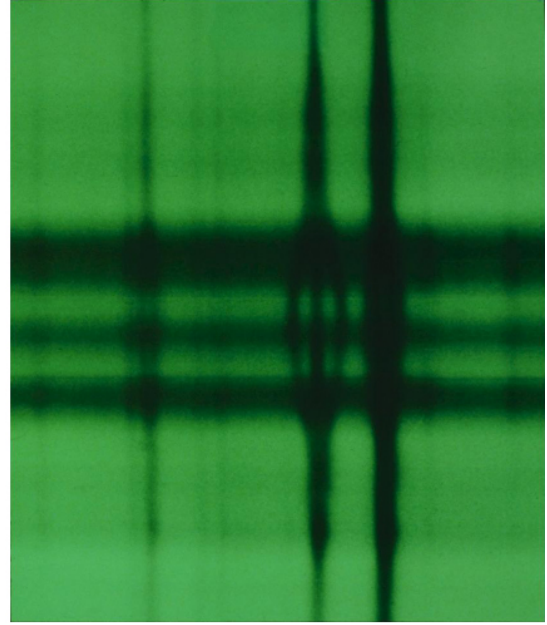
In the presence of a strong magnetic field, however, each energy level is separated into several levels very close to one another. The separation of the levels is proportional to the strength of the field. As a result, spectral lines formed in the presence of a magnetic field are not single lines but a series of very closely spaced lines corresponding to the subdivisions of the atomic energy levels. This splitting of lines in the presence of a magnetic field is what we call the Zeeman effect (after the Dutch scientist who first discovered it in 1896).

Measurements of the Zeeman effect in the spectra of the light from sunspot regions show them to have strong magnetic fields ([link](#)). Bear in mind that magnets always have a north pole and a south pole. Whenever sunspots are observed in pairs, or in groups containing two principal spots, one of the spots usually has the magnetic polarity of a north-seeking magnetic pole and the other has the opposite polarity. Moreover, during a given cycle, the leading spots of pairs (or leading principle spots of groups) in the Northern Hemisphere all tend to have the same polarity, whereas those in the Southern Hemisphere all tend to have the opposite polarity.

Zeeman Effect.



(a)

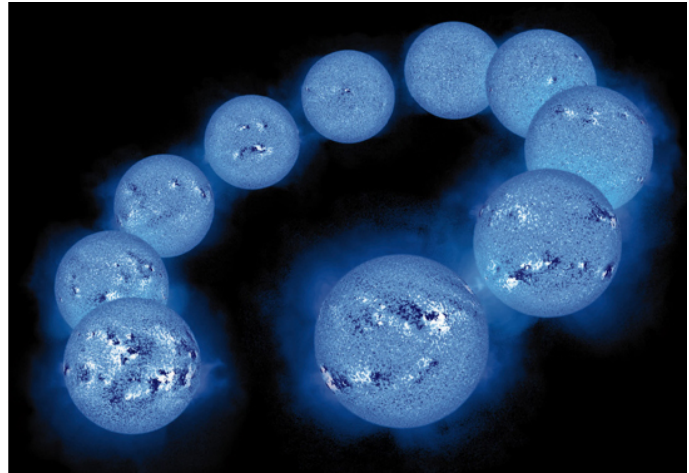
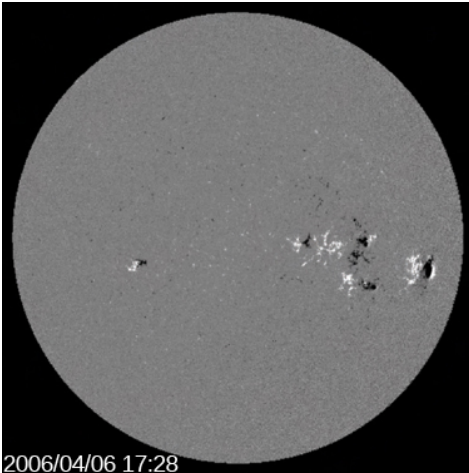


(b)

These photographs show how magnetic fields in sunspots are measured by means of the Zeeman effect. (left) The vertical black line indicates the position of the spectrograph slit through which light is passed to obtain the spectrum in (right). (credit: modification of work by NSO/AURA/NSF)

During the next sunspot cycle, however, the polarity of the leading spots is reversed in each hemisphere. For example, if during one cycle, the leading spots in the Northern Hemisphere all had the polarity of a north-seeking pole, then the leading spots in the Southern Hemisphere would have the polarity of a south-seeking pole. During the next cycle, the leading spots in the Northern Hemisphere would have south-seeking polarity, whereas those in the Southern Hemisphere would have north-seeking polarity. Therefore, strictly speaking, the sunspot cycle does not repeat itself in regard to magnetic polarity until two 11-year cycles have passed. A visual representation of the Sun's magnetic fields, called a *magnetogram*, can be used to see the relationship between sunspots and the Sun's magnetic field ([link](#)).

Magnetogram and Solar Cycle.



In the image on the left, called a magnetogram, we see the magnetic polarity of sunspots. The black areas are where the magnetism is pointing toward the Sun's core, whereas the white regions are where it is pointing away from the core, toward us. This dramatic sequence on the right shows the activity cycle of the Sun. The 10 maps of the magnetic field on the surface of the Sun span a period of 7.5 years. The two magnetic polarities (N and S) of the magnetic field are shown against a blue disk as dark blue to black (N) and as light blue to white (S). The earliest image, taken on January 8, 1992, is at the lower left and was taken just after solar maximum. Each image, from left to right around the arc, was taken one-half to one year after the preceding one.

The last image was taken on July 25, 1999, as the Sun was approaching the next solar maximum. Note a few striking patterns in the magnetic maps: the direction from white to black polarity in the Southern Hemisphere is opposite from that in the Northern Hemisphere. (credit left: modification of work by NASA/SDO; credit right: modification of work by NASA/SOHO)

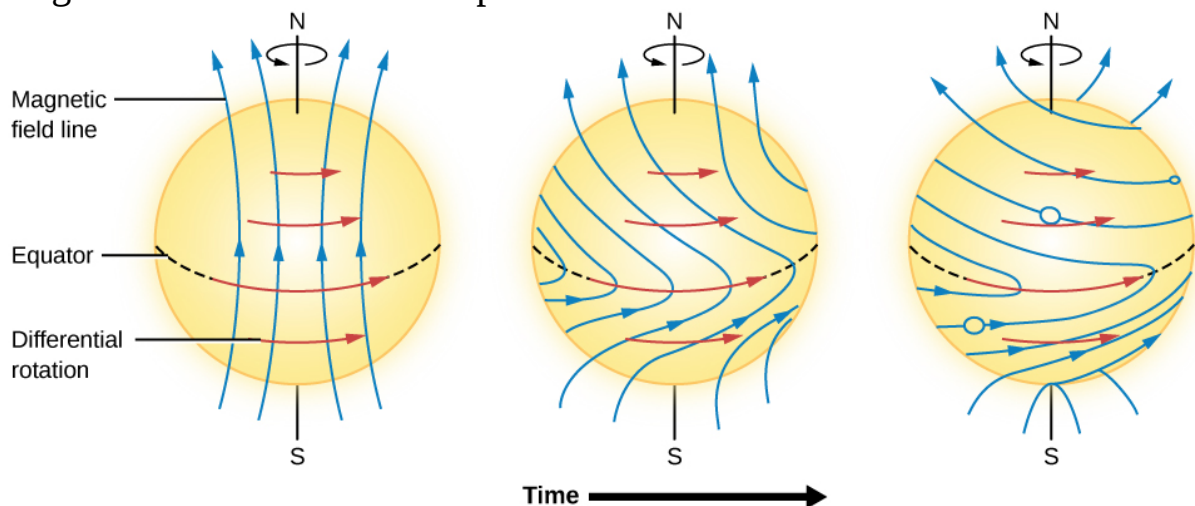
Why is the Sun such a strong and complicated magnet? Astronomers have found that it is the Sun's *dynamo* that generates the magnetic field. A dynamo is a machine that converts kinetic energy (i.e., the energy of motion) into electricity. On Earth, dynamos are found in power plants where, for example, the energy from wind or flowing water is used to cause turbines to rotate. In the Sun, the source of kinetic energy is the churning of

turbulent layers of ionized gas within the Sun's interior that we mentioned earlier. These generate electric currents—moving electrons—which in turn generate magnetic fields.

Most solar researchers agree that the solar dynamo is located in the convection zone or in the interface layer between the convection zone and the radiative zone below it. As the magnetic fields from the Sun's dynamo interact, they break, reconnect, and rise through the Sun's surface.

We should say that, although we have good observations that show us *how* the Sun changes during each solar cycle, it is still very difficult to build physical models of something as complicated as the Sun that can account satisfactorily for *why* it changes. Researchers have not yet developed a generally accepted model that describes in detail the physical processes that control the solar cycle. Calculations do show that differential rotation (the idea that the Sun rotates at different rates at different latitudes) and convection just below the solar surface can twist and distort the magnetic fields. This causes them to grow and then decay, regenerating with opposite polarity approximately every 11 years. The calculations also show that as the fields grow stronger near solar maximum, they flow from the interior of the Sun toward its surface in the form of loops. When a large loop emerges from the solar surface, it creates regions of sunspot activity ([\[link\]](#)).

Magnetic Field Lines Wind Up.



Because the Sun spins faster at the equator than near the poles, the magnetic fields in the Sun tend to wind up as shown, and after a while

make loops. This is an idealized diagram; the real situation is much more complex.

This idea of magnetic loops offers a natural explanation of why the leading and trailing sunspots in an active region have opposite polarity. The leading sunspot coincides with one end of the loop and the trailing spot with the other end. Magnetic fields also hold the key to explaining why sunspots are cooler and darker than the regions without strong magnetic fields. The forces produced by the magnetic field resist the motions of the bubbling columns of rising hot gases. Since these columns carry most of the heat from inside the Sun to the surface by means of convection, and strong magnetic fields inhibit this convection, the surface of the Sun is allowed to cool. As a result, these regions are seen as darker, cooler sunspots.

Beyond this general picture, researchers are still trying to determine why the magnetic fields are as large as they are, why the polarity of the field in each hemisphere flips from one cycle to the next, why the length of the solar cycle can vary from one cycle to the next, and why events like the Maunder Minimum occur.

Note:

In this [video](#) solar scientist Holly Gilbert discusses the Sun's magnetic field.

Key Concepts and Summary

Sunspots are dark regions where the temperature is up to 2000 K cooler than the surrounding photosphere. Their motion across the Sun's disk allows us to calculate how fast the Sun turns on its axis. The Sun rotates more rapidly at its equator, where the rotation period is about 25 days, than near the poles, where the period is slightly longer than 36 days. The number of visible sunspots varies according to a sunspot cycle that averages 11

years in length. Spots frequently occur in pairs. During a given 11-year cycle, all leading spots in the Northern Hemisphere have the same magnetic polarity, whereas all leading spots in the Southern Hemisphere have the opposite polarity. In the subsequent 11-year cycle, the polarity reverses. For this reason, the magnetic activity cycle of the Sun is understood to last for 22 years. This activity cycle is connected with the behavior of the Sun's magnetic field, but the exact mechanism is not yet understood.

Glossary

differential rotation

the phenomenon that occurs when different parts of a rotating object rotate at different rates at different latitudes

Maunder Minimum

a period during the seventeenth century when the number of sunspots seen throughout the solar cycle was unusually low

sunspot

large, dark features seen on the surface of the Sun caused by increased magnetic activity

sunspot cycle

the semiregular 11-year period with which the frequency of sunspots fluctuates

Solar Activity above the Photosphere

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the various ways in which the solar activity cycle manifests itself, including flares, coronal mass ejections, prominences, and plages

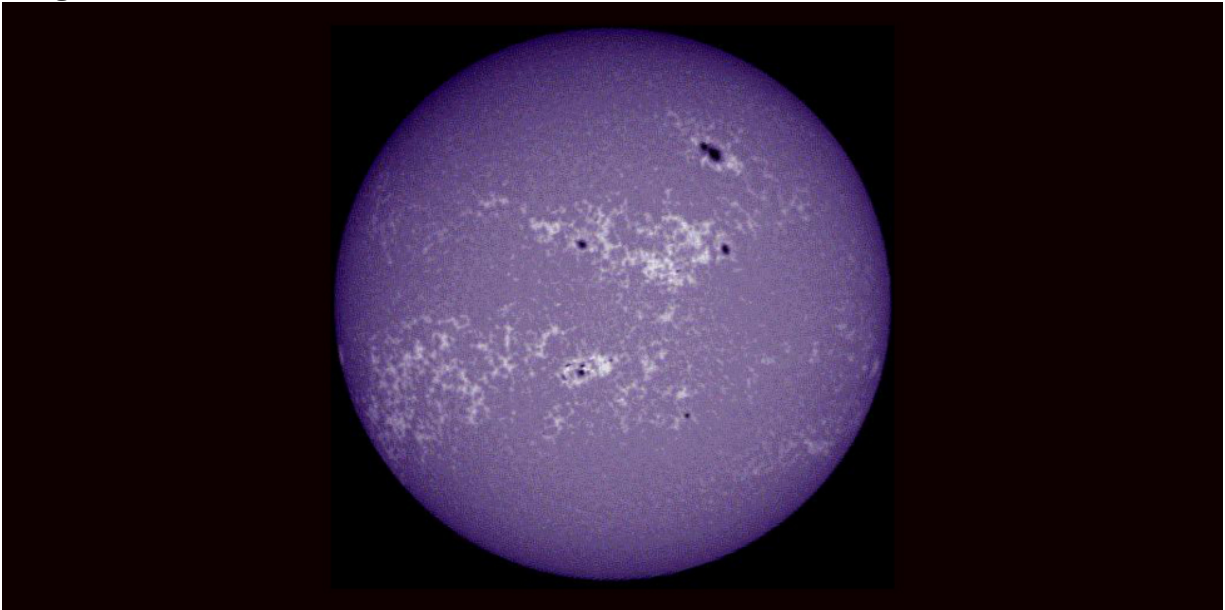
Sunspots are not the only features that vary during a solar cycle. There are dramatic changes in the chromosphere and corona as well. To see what happens in the chromosphere, we must observe the emission lines from elements such as hydrogen and calcium, which emit useful spectral lines at the temperatures in that layer. The hot corona, on the other hand, can be studied by observations of X-rays and of extreme ultraviolet and other wavelengths at high energies.

Plages and Prominences

As we saw, emission lines of hydrogen and calcium are produced in the hot gases of the chromosphere. Astronomers routinely photograph the Sun through filters that transmit light only at the wavelengths that correspond to these emission lines. Pictures taken through these special filters show bright “clouds” in the chromosphere around sunspots; these bright regions are known as **plages** ([\[link\]](#)). These are regions within the chromosphere that have higher temperature and density than their surroundings. The plages actually contain all of the elements in the Sun, not just hydrogen and

calcium. It just happens that the spectral lines of hydrogen and calcium produced by these clouds are bright and easy to observe.

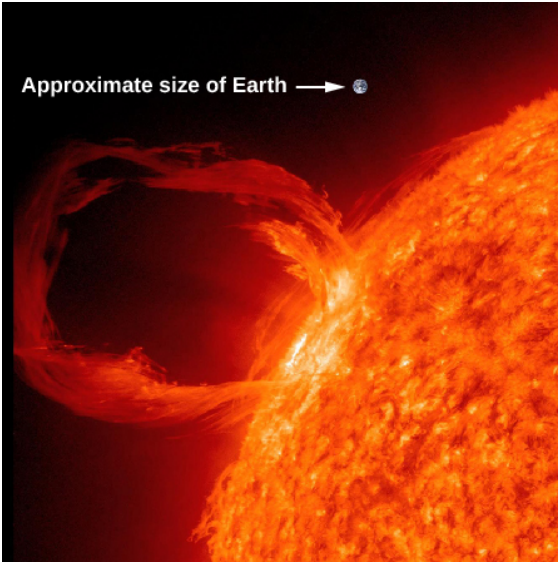
Plages on the Sun.



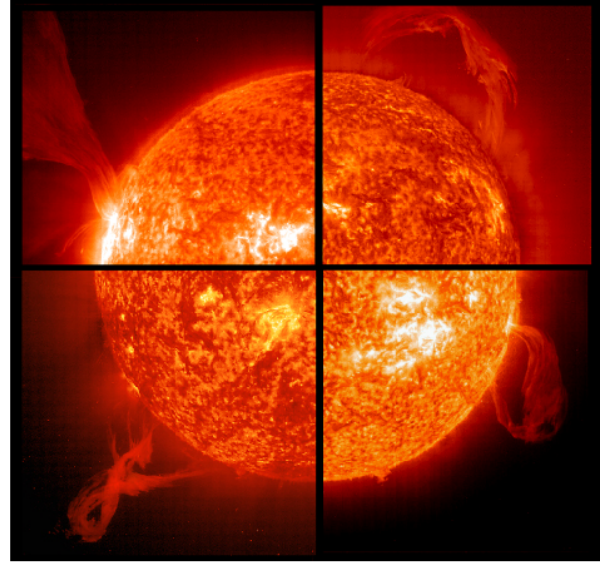
This image of the Sun was taken with a filter that transmits only the light of the spectral line produced by singly ionized calcium. The bright cloud-like regions are the plages. (credit: modification of work by NASA)

Moving higher into the Sun's atmosphere, we come to the spectacular phenomena called **prominences** ([\[link\]](#)), which usually originate near sunspots. Eclipse observers often see prominences as red features rising above the eclipsed Sun and reaching high into the corona. Some, the *quiescent* prominences, are graceful loops of plasma (ionized gas) that can remain nearly stable for many hours or even days. The relatively rare *eruptive* prominences appear to send matter upward into the corona at high speeds, and the most active *surge* prominences may move as fast as 1300 kilometers per second (almost 3 million miles per hour). Some eruptive prominences have reached heights of more than 1 million kilometers above the photosphere; Earth would be completely lost inside one of those awesome displays ([\[link\]](#)).

Prominences.



(a)



(b)

(a) This image of an eruptive prominence was taken in the light of singly ionized helium in the extreme ultraviolet part of the spectrum. The prominence is a particularly large one. An image of Earth is shown at the same scale for comparison.

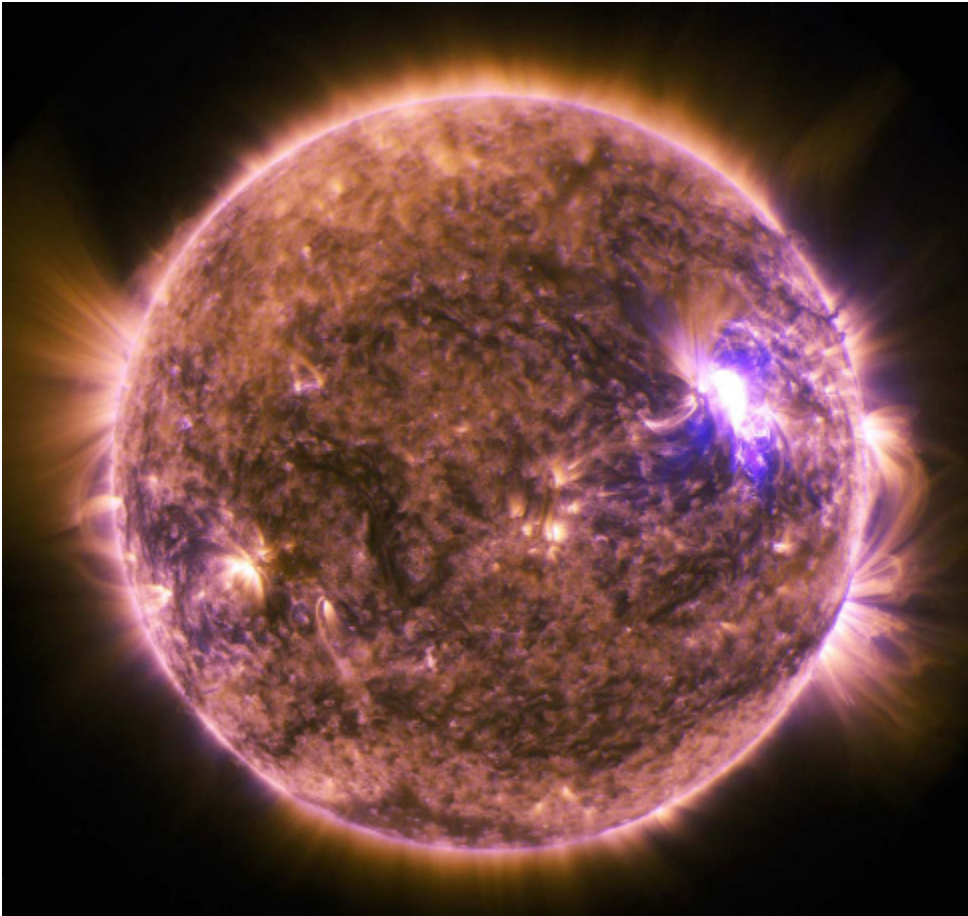
(b) A prominence is a huge cloud of relatively cool (about 60,000 K in this case), fairly dense gas suspended in the much hotter corona. These pictures, taken in ultraviolet, are color coded so that white corresponds to the hottest temperatures and dark red to cooler ones. The four images were taken, moving clockwise from the upper left, on May 15, 2001; March 28, 2000; January 18, 2000; and February 2, 2001. (credit a: modification of work by NASA/SOHO; credit b: modification of work by NASA/SDO)

Flares and Coronal Mass Ejections

The most violent event on the surface of the Sun is a rapid eruption called a **solar flare** ([link](#)). A typical flare lasts for 5 to 10 minutes and releases a total amount of energy equivalent to that of perhaps a million hydrogen bombs. The largest flares last for several hours and emit enough energy to power the entire United States at its current rate of electrical consumption

for 100,000 years. Near sunspot maximum, small flares occur several times per day, and major ones may occur every few weeks.

Solar Flare.



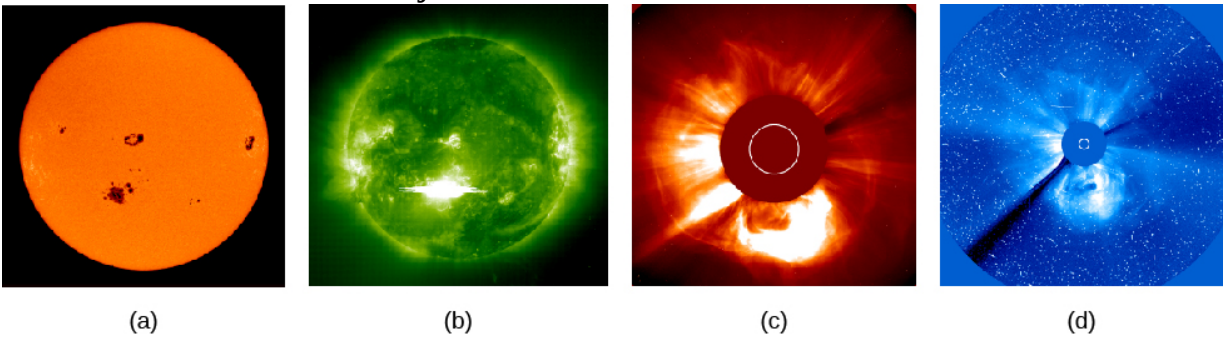
The bright white area seen on the right side of the Sun in this image from the Solar Dynamics Observer spacecraft is a solar flare that was observed on June 25, 2015. (credit: NASA/SDO)

Flares, like the one shown in [\[link\]](#), are often observed in the red light of hydrogen, but the visible emission is only a tiny fraction of the energy released when a solar flare explodes. At the moment of the explosion, the matter associated with the flare is heated to temperatures as high as 10 million K. At such high temperatures, a flood of X-ray and ultraviolet radiation is emitted.

Flares seem to occur when magnetic fields pointing in opposite directions release energy by interacting with and destroying each other—much as a stretched rubber band releases energy when it breaks.

What is different about flares is that their magnetic interactions cover a large volume in the solar corona and release a tremendous amount of electromagnetic radiation. In some cases, immense quantities of coronal material—mainly protons and electrons—may also be ejected at high speeds (500–1000 kilometers per second) into interplanetary space. Such a **coronal mass ejection (CME)** can affect Earth in a number of ways (which we will discuss in the section on space weather).

Flare and Coronal Mass Ejection.



This sequence of four images shows the evolution over time of a giant eruption on the Sun. (a) The event began at the location of a sunspot group, and (b) a flare is seen in far-ultraviolet light. (c) Fourteen hours later, a CME is seen blasting out into space. (d) Three hours later, this CME has expanded to form a giant cloud of particles escaping from the Sun and is beginning the journey out into the solar system. The white circle in (c) and (d) shows the diameter of the solar photosphere. The larger dark area shows where light from the Sun has been blocked out by a specially designed instrument to make it possible to see the faint emission from the corona. (credit a, b, c, d: modification of work by SOHO/EIT, SOHO/LASCO, SOHO/MDI (ESA & NASA))

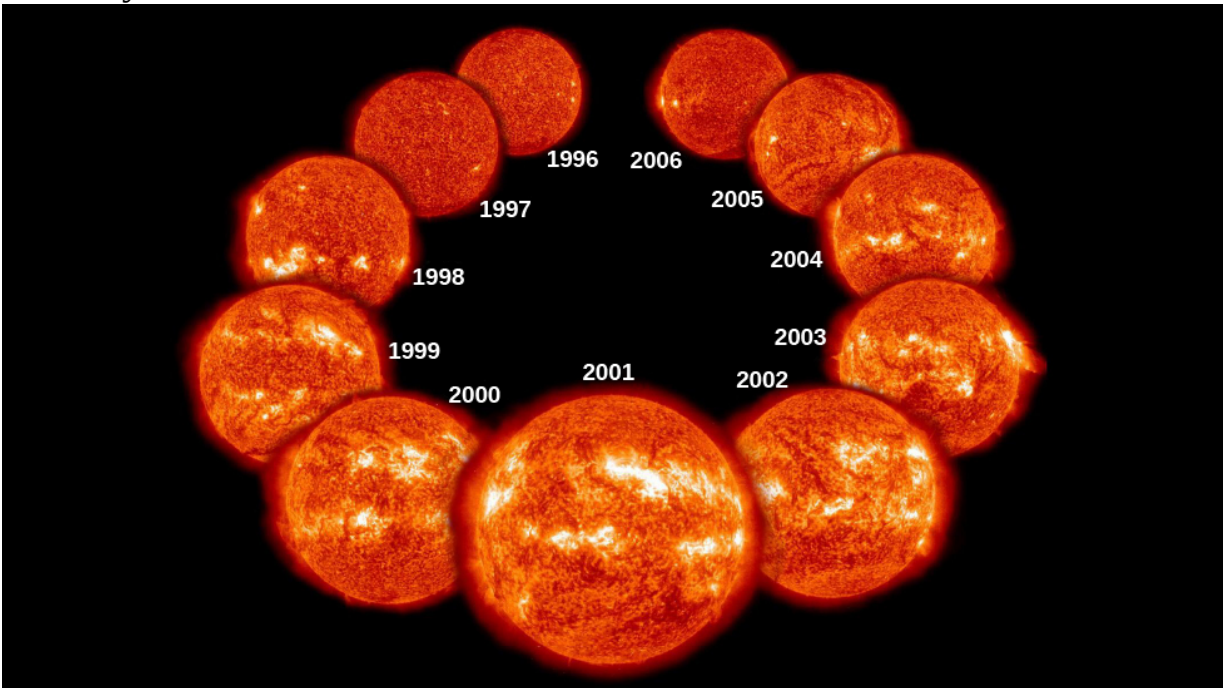
Note:

See a [coronal mass ejection](#) recorded by the Solar Dynamics Observatory.

Active Regions

To bring the discussion of the last two sections together, astronomers now realize that sunspots, flares, and bright regions in the chromosphere and corona tend to occur together on the Sun in time and space. That is, they all tend to have similar longitudes and latitudes, but they are located at different heights in the atmosphere. Because they all occur together, they vary with the sunspot cycle.

Solar Cycle.

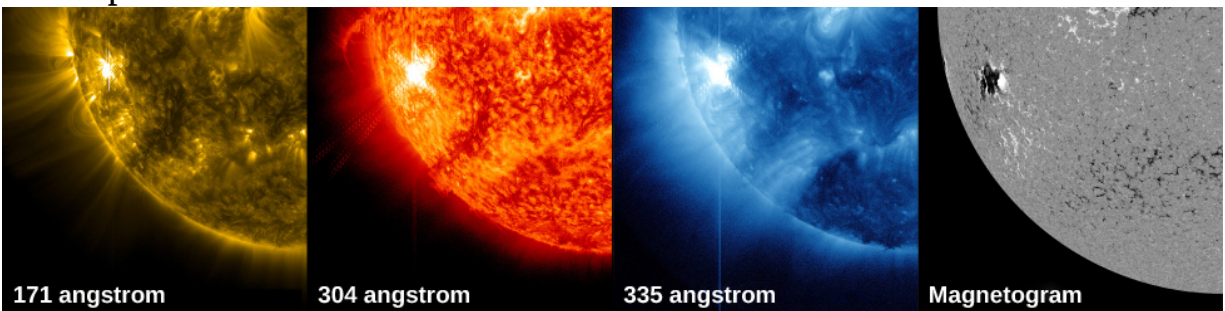


This dramatic sequence of images taken from the SOHO satellite over a period of 11 years shows how active regions change during the solar cycle. The images were taken in the ultraviolet region of the spectrum and show that active regions on the Sun increase and decrease during the cycle. Sunspots are located in the cooler photosphere, beneath the hot gases shown in this image, and vary in phase with the emission from these hot gases—more sunspots and more emission from hot

gases occur together. (credit: modification of work by
ESA/NASA/SOHO)

For example, flares are more likely to occur near sunspot maximum, and the corona is much more conspicuous at that time (see [\[link\]](#)). A place on the Sun where a number of these phenomena are seen is called an **active region** ([\[link\]](#)). As you might deduce from our earlier discussion, active regions are always associated with strong magnetic fields.

Solar Active Region Observed at Different Heights in the Sun's Atmosphere.



These four images of a solar flare on October 22, 2012, show from the left: light from the Sun at a wavelength of 171 angstroms, which shows the structure of loops of solar material in the corona; ultraviolet at 304 angstroms, which shows light from the region of the Sun's atmosphere where flares originate; light at 335 angstroms, which highlights radiation from active regions in the corona; a magnetogram, which shows magnetically active regions on the Sun. Note how these different types of activity all occur above a sunspot region with a strong magnetic field. (credit: modification of work by
NASA/SDO/Goddard)

Key Concepts and Summary

Signs of more intense solar activity, an increase in the number of sunspots, as well as prominences, plagues, solar flares, and coronal mass ejections, all tend to occur in active regions—that is, in places on the Sun with the same

latitude and longitude but at different heights in the atmosphere. Active regions vary with the solar cycle, just like sunspots do.

Glossary

active region

an area on the Sun where magnetic fields are concentrated; sunspots, prominences, flares, and CMEs all tend to occur in active regions

coronal mass ejection (CME)

a solar flare in which immense quantities of coronal material—mainly protons and electrons—is ejected at high speeds (500–1000 kilometers per second) into interplanetary space

plage

a bright region of the solar surface observed in the light of some spectral line

prominence

a large, bright, gaseous feature that appears above the surface of the Sun and extends into the corona

solar flare

a sudden and temporary outburst of electromagnetic radiation from an extended region of the Sun's surface

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain what space weather is and how it affects Earth

In the previous sections, we have seen that some of the particles coming off the Sun—either steadily as in the solar wind or in great bursts like CMEs—will reach Earth and its *magnetosphere* (the zone of magnetic influence that surrounds our planet). As if scientists did not have enough trouble trying to predict weather on Earth, this means that they are now facing the challenge of predicting the effects of solar storms on Earth. This field of research is called *space weather*; when that weather turns stormy, our technology turns out to be at risk.

With thousands of satellites in orbit, astronauts taking up long-term residence in the International Space Station, millions of people using cell phones, GPS, and wireless communication, and nearly everyone relying on the availability of dependable electrical power, governments are now making major investments in trying to learn how to predict when solar storms will occur and how strongly they will affect Earth.

Some History

What we now study as space weather was first recognized (though not yet understood) in 1859, in what is now known as the Carrington Event. In early September of that year, two amateur astronomers, including Richard Carrington in England, independently observed a solar flare. This was followed a day or two later by a significant solar storm reaching the region of Earth's magnetic field, which was soon overloaded with charged particles (see [Earth as a Planet](#)).

As a result, aurora activity was intense and the northern lights were visible well beyond their normal locations near the poles—as far south as Hawaii and the

Caribbean. The glowing lights in the sky were so intense that some people reported getting up in the middle of the night, thinking it must be daylight.

The 1859 solar storm happened at a time when a new technology was beginning to tie people in the United States and some other countries together: the telegraph system. This was a machine and network for sending messages in code through overhead electrical wires (a bit like a very early version of the internet). The charged particles that overwhelmed Earth's magnetic field descended toward our planet's surface and affected the wires of the telegraph system. Sparks were seen coming out of exposed wires and out of the telegraph machines in the system's offices.

The observation of the bright flare that preceded these effects on Earth led to scientific speculation that a connection existed between solar activity and impacts on Earth—this was the beginning of our understanding of what today we call space weather.

Note:

Watch NASA scientists [answer some questions](#) about space weather, and [discuss](#) some effects it can have in space and on Earth.

Sources of Space Weather

Three solar phenomena—coronal holes, solar flares, and CMEs—account for most of the space weather we experience. Coronal holes allow the solar wind to flow freely away from the Sun, unhindered by solar magnetic fields. When the solar wind reaches Earth, as we saw, it causes Earth's magnetosphere to contract and then expand after the solar wind passes by. These changes can cause (usually mild) electromagnetic disturbances on Earth.

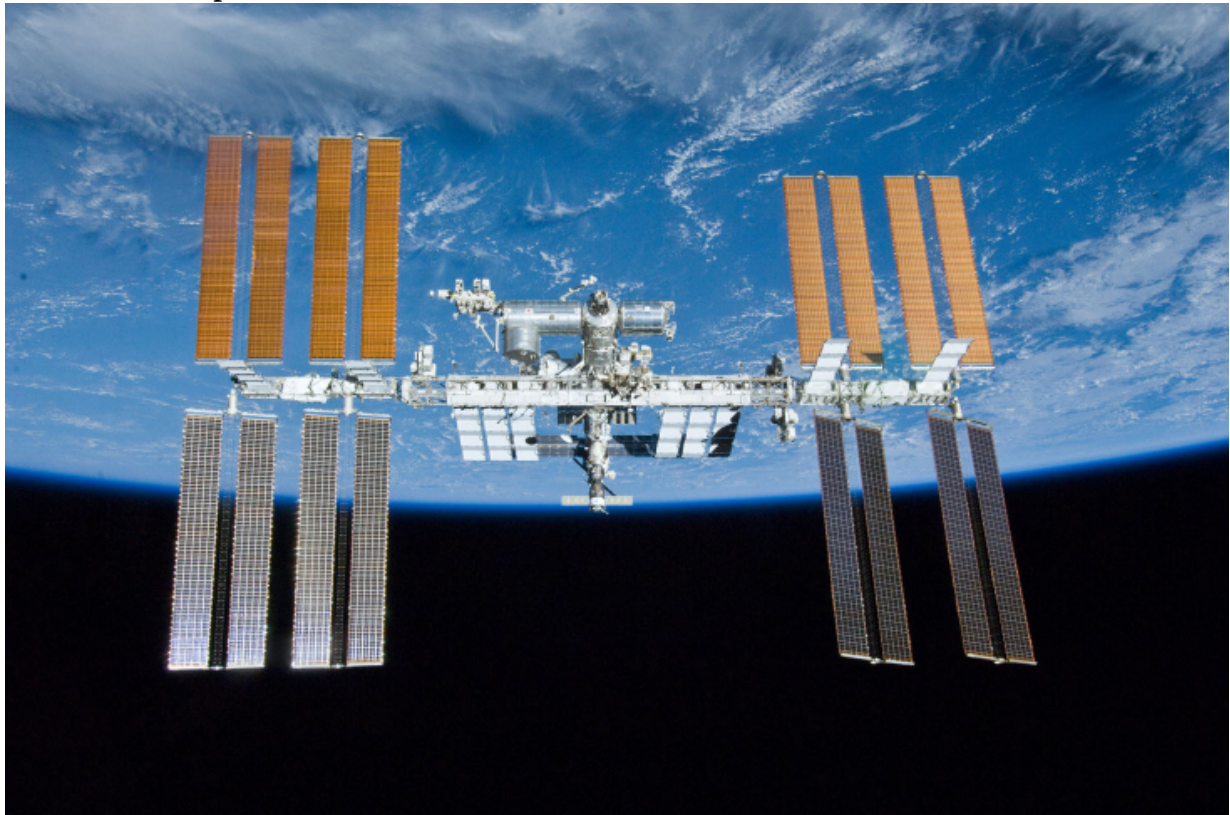
More serious are solar flares, which shower the upper atmosphere of Earth with X-rays, energetic particles, and intense ultraviolet radiation. The X-rays and ultraviolet radiation can ionize atoms in Earth's upper atmosphere, and the freed electrons can build up a charge on the surface of a spacecraft. When this static charge discharges, it can damage the electronics in the spacecraft—just as you can receive a shock when you walk across a carpet in your stocking feet in a dry climate and then touch a light switch or some other metal object.

Most disruptive are coronal mass ejections. A CME is an erupting bubble of tens of millions of tons of gas blown away from the Sun into space. When this bubble reaches Earth a few days after leaving the Sun, it heats the ionosphere, which expands and

reaches farther into space. As a consequence, friction between the atmosphere and spacecraft increases, dragging satellites to lower altitudes.

At the time of a particularly strong flare and CME in March 1989, the system responsible for tracking some 19,000 objects orbiting Earth temporarily lost track of 11,000 of them because their orbits were changed by the expansion of Earth's atmosphere. During solar maximum, a number of satellites are brought to such a low altitude that they are destroyed by friction with the atmosphere. Both the Hubble Space Telescope and the International Space Station ([link](#)) require reboosts to higher altitude so that they can remain in orbit.

International Space Station.



The International Space Station is seen above Earth, as photographed in 2010 by the departing crew of the Space Shuttle Atlantis. (credit: NASA)

Solar Storm Damage on Earth

When a CME reaches Earth, it distorts Earth's magnetic field. Since a changing magnetic field induces electrical current, the CME accelerates electrons, sometimes to very high speeds. These "killer electrons" can penetrate deep into satellites,

sometimes destroying their electronics and permanently disabling operation. This has happened with some communications satellites.

Disturbances in Earth's magnetic field can cause disruptions in communications, especially cell phone and wireless systems. In fact, disruptions can be expected to occur several times a year during solar maximum. Changes in Earth's magnetic field due to CMEs can also cause surges in power lines large enough to burn out transformers and cause major power outages. For example, in 1989, parts of Montreal and Quebec Province in Canada were without power for up to 9 hours as a result of a major solar storm. Electrical outages due to CMEs are more likely to occur in North America than in Europe because North America is closer to Earth's magnetic pole, where the currents induced by CMEs are strongest.

Besides changing the orbits of satellites, CMEs can also distort the signals sent by them. These effects can be large enough to reduce the accuracy of GPS-derived positions so that they cannot meet the limits required for airplane systems, which must know their positions to within 160 feet. Such disruptions caused by CMEs have occasionally forced the Federal Aviation Administration to restrict flights for minutes or, in a few cases, even days.

Solar storms also expose astronauts, passengers in high-flying airplanes, and even people on the surface of Earth to increased amounts of radiation. Astronauts, for example, are limited in the total amount of radiation to which they can be exposed during their careers. A single ill-timed solar outburst could end an astronaut's career. This problem becomes increasingly serious as astronauts spend more time in space. For example, the typical *daily* dose of radiation aboard the Russian Mir space station was equivalent to about eight chest X-rays. One of the major challenges in planning the human exploration of Mars is devising a way to protect astronauts from high-energy solar radiation.

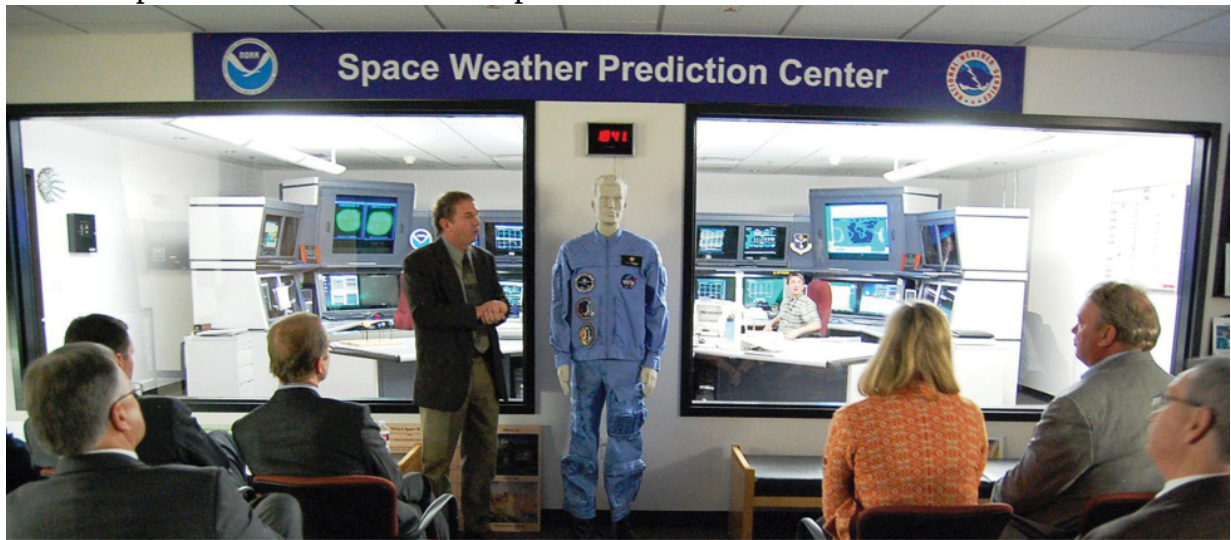
Advance warning of solar storms would help us minimize their disruptive effects. Power networks could be run at less than their full capacity so that they could absorb the effects of power surges. Communications networks could be prepared for malfunctions and have backup plans in place. Spacewalks could be timed to avoid major solar outbursts. Scientists are now trying to find ways to predict where and when flares and CMEs will occur, and whether they will be big, fast events or small, slow ones with little consequence for Earth.

The strategy is to relate changes in the appearance of small, active regions and changes in local magnetic fields on the Sun to subsequent eruptions. However, right now, our predictive capability is still poor, and so the only real warning we have is from actually seeing CMEs and flares occur. Since a CME travels outward at about 500 kilometers per second, an observation of an eruption provides several days

warning at the distance of Earth. However, the severity of the impact on Earth depends on how the magnetic field associated with the CME is oriented relative to Earth's magnetic field. The orientation can be measured only when the CME flows past a satellite we have put up for this purpose. However, it is located only about an hour upstream from Earth.

Space weather predictions are now available online to scientists and the public. Outlooks are given a week ahead, bulletins are issued when there is an event that is likely to be of interest to the public, and warnings and alerts are posted when an event is imminent or already under way ([link](#)).

NOAA Space Weather Prediction Operations Center.



Bill Murtagh, a space weather forecaster, leads a workshop on preparedness for events like geomagnetic storms. (credit: modification of work by FEMA/Jerry DeFelice)

Note:

To find public information and alerts about space weather, you can turn to the [National Space Weather Prediction Center](#) or [SpaceWeather](#) for consolidated information from many sources.

Fortunately, we can expect calmer space weather for the next few years, since the most recent solar maximum, which was relatively weak, occurred in 2014, and

scientists believe the current solar cycle to be one of the least active in recent history. We expect more satellites to be launched that will allow us to determine whether CMEs are headed toward Earth and how big they are. Models are being developed that will then allow scientists to use early information about the CME to predict its likely impact on Earth.

The hope is that by the time of the next maximum, solar weather forecasting will have some of the predictive capability that meteorologists have achieved for terrestrial weather at Earth's surface. However, the most difficult events to predict are the largest and most damaging storms—hurricanes on Earth and extreme, rare storm events on the Sun. Thus, it is inevitable that the Sun will continue to surprise us.

Example:**The Timing of Solar Events**

A basic equation is useful in figuring out when events on the Sun will impact Earth:

Equation:

$$\text{distance} = \text{velocity} \times \text{time, or } D = v \times t$$

Dividing both sides by v , we get

Equation:

$$T = D/v$$

Suppose you observe a major solar flare while astronauts are orbiting Earth. If the average speed of solar wind is 400 km/s and the distance to the Sun as 1.496×10^8 km, how long it will before the charged particles ejected from the Sun during the flare reach the space station?

Solution

The time required for solar wind particles to reach Earth is $T = D/v$.

Equation:

$$\frac{1.496 \times 10^8 \text{ km}}{400 \text{ km/s}} = 3.74 \times 10^5 \text{ s, or } \frac{3.74 \times 10^5 \text{ s}}{60 \text{ s/min} \times 60 \text{ min/h} \times 24 \text{ h/d}} = 4.3 \text{ d}$$

Check Your Learning

How many days would it take for the particles to reach Earth if the solar wind speed increased to 500 km/s?

Note:

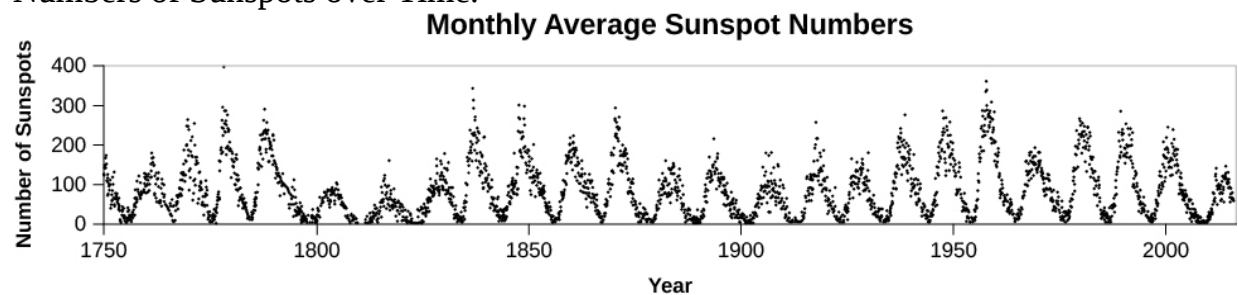
Answer:

$$\frac{1.496 \times 10^8 \text{ km}}{500 \text{ km/s}} = 2.99 \times 10^5 \text{ s, or } \frac{2.99 \times 10^5 \text{ s}}{60 \text{ s/min} \times 60 \text{ min/h} \times 24 \text{ h/d}} = 3.46 \text{ d}$$

Earth's Climate and the Sunspot Cycle: Is There a Connection?

While the Sun rises faithfully every day at a time that can be calculated precisely, scientists have determined that the Sun's energy output is not truly constant but varies over the centuries by a small amount—probably less than 1%. We've seen that the number of sunspots varies, with the time between sunspot maxima of about 11 years, and that the number of sunspots at maximum is not always the same. Considerable evidence shows that between the years 1645 and 1715, the number of sunspots, even at sunspot maximum, was much lower than it is now. This interval of significantly low sunspot numbers was first noted by Gustav Spörer in 1887 and then by E. W. Maunder in 1890; it is now called the **Maunder Minimum**. The variation in the number of sunspots over the past three centuries is shown in [\[link\]](#). Besides the Maunder Minimum in the seventeenth century, sunspot numbers were somewhat lower during the first part of the nineteenth century than they are now; this period is called the Little Maunder Minimum.

Numbers of Sunspots over Time.



This diagram shows how the number of sunspots has changed with time since counts of the numbers of spots began to be recorded on a consistent scale. Note the low number of spots during the early years of the nineteenth century, the Little Maunder Minimum. (credit: modification of work by NASA/ARC)

When the number of sunspots is high, the Sun is active in various other ways as well, and, as we will see in several sections below, some of this activity affects Earth

directly. For example, there are more auroral displays when the sunspot number is high. Auroras are caused when energetically charged particles from the Sun interact with Earth's magnetosphere, and the Sun is more likely to spew out particles when it is active and the sunspot number is high. Historical accounts also indicate that auroral activity was abnormally low throughout the several decades of the Maunder Minimum.

The Maunder Minimum was a time of exceptionally low temperatures in Europe—so low that this period is described as the Little Ice Age. This coincidence in time caused scientists to try to understand whether small changes in the Sun could affect the climate on Earth. There is clear evidence that it was unusually cold in Europe during part of the seventeenth century. The River Thames in London froze at least 11 times, ice appeared in the oceans off the coasts of southeast England, and low summer temperatures led to short growing seasons and poor harvests. However, whether and how changes on the Sun on this timescale influence Earth's climate is still a matter of debate among scientists.

Other small changes in climate like the Little Ice Age have occurred and have had their impacts on human history. For example, explorers from Norway first colonized Iceland and then reached Greenland by 986. From there, they were able to make repeated visits to the northeastern coasts of North America, including Newfoundland, between about 1000 and 1350. (The ships of the time did not allow the Norse explorers to travel all the way to North America directly, but only from Greenland, which served as a station for further exploration.)

Most of Greenland is covered by ice, and the Greenland station was never self-sufficient; rather, it depended on imports of food and other goods from Norway for its survival. When a little ice age began in the thirteenth century, voyaging became very difficult, and support of the Greenland colony was no longer possible. The last-known contact with it was made by a ship from Iceland blown off course in 1410. When European ships again began to visit Greenland in 1577, the entire colony there had disappeared.

The estimated dates for these patterns of migration follow what we know about solar activity. Solar activity was unusually high between 1100 and 1250, which includes the time when the first European contacts were made with North America. Activity was low from 1280 to 1340 and there was a little ice age, which was about the time regular contact with North America and between Greenland and Europe stopped.

One must be cautious, however, about assuming that low sunspot numbers or variations in the Sun's output of energy *caused* the Little Ice Age. There is no satisfactory model that can explain how a reduction in solar activity might cause cooler temperatures on Earth. An alternative possibility is that the cold weather during

the Little Ice Age was related to volcanic activity. Volcanoes can eject aerosols (tiny droplets or particles) into the atmosphere that efficiently reflect sunlight. Observations show, for example, that the Pinatubo eruption in 1991 ejected SO₂ aerosols into the atmosphere, which reduced the amount of sunlight reaching Earth's surface enough to lower global temperatures by 0.4 °C.

Satellite data show that the energy output from the Sun during a solar cycle varies by only about 0.1%. We know of no physical process that would explain how such a small variation could cause global temperature changes. The level of solar activity may, however, have other effects. For example, although the Sun's total energy output varies by only 0.1% during a solar cycle, its extreme ultraviolet radiation is 10 times higher at times of solar maximum than at solar minimum. This large variation can affect the chemistry and temperature structure of the upper atmosphere. One effect might be a reduction in the ozone layer and a cooling of the stratosphere near Earth's poles. This, in turn, could change the circulation patterns of winds aloft and, hence, the tracks of storms. There is some recent evidence that variations in regional rainfall correlate better with solar activity than does the global temperature of Earth. But, as you can see, the relationship between what happens on the Sun and what happens to Earth's climate over the short term is still an area that scientists are investigating and debating.

Whatever the effects of solar activity may be on local rainfall or temperature patterns, we want to emphasize one important idea: Our climate change data and the models developed to account for the data consistently show that solar variability is *not* the cause of the global warming that has occurred during the past 50 years.

Key Concepts and Summary

Space weather is the effect of solar activity on our own planet, both in our magnetosphere and on Earth's surface. Coronal holes allow more of the Sun's material to flow out into space. Solar flares and coronal mass ejections can cause auroras, disrupt communications, damage satellites, and cause power outages on Earth.

For Further Exploration

Articles

Berman, B. "How Solar Storms Could Shut Down Earth." *Astronomy* (September 2013): 22. Up-to-date review of how events on the Sun can hurt our civilization.

Frank, A. “Blowin’ in the Solar Wind.” *Astronomy* (October 1998): 60. On results from the SOHO spacecraft.

Holman, G. “The Mysterious Origins of Solar Flares.” *Scientific American* (April 2006): 38. New ideas involving magnetic reconnection and new observations of flares.

James, C. “Solar Forecast: Storm Ahead.” *Sky & Telescope* (July 2007): 24. Nice review on the effects of the Sun’s outbursts and on Earth and how we monitor “space weather.”

Schaefer, B. “Sunspots That Changed the World.” *Sky & Telescope* (April 1997): 34. Historical events connected with sunspots and solar activity.

Schrijver, C. and Title, A. “Today’s Science of the Sun.” *Sky & Telescope* (February 2001): 34; (March 2001): 34. Excellent reviews of recent results about the solar atmosphere.

Wadhwa, M. “Order from Chaos: Genesis Samples the Solar Wind.” *Astronomy* (October 2013): 54. On a satellite that returned samples of the Sun’s wind.

Websites

Dr. Sten Odenwald’s “Solar Storms” site: <http://www.solarstorms.org/>.

ESA/NASA’s Solar & Heliospheric Observatory: <http://sohowww.nascom.nasa.gov>. A satellite mission with a rich website to explore.

High Altitude Observatory Introduction to the Sun:
<http://www.hao.ucar.edu/education/basic.php>. For beginners.

NASA’s Solar Missions:
https://www.nasa.gov/mission_pages/sunearth/missions/index.html. Good summary of the many satellites and missions NASA has.

NOAA Profile of Space Weather:
http://www.swpc.noaa.gov/sites/default/files/images/u33/primer_2010_new.pdf. A primer.

NOAA Space Weather Prediction Center Information Pages:
<http://www.swpc.noaa.gov/content/education-and-outreach>. Includes primers, videos, a curriculum and training modules.

Nova Sun Lab: <http://www.pbs.org/wgbh/nova/labs/lab/sun/>. Videos, scientist profiles, a research challenge related to the active Sun from the PBS science program.

Space Weather: Storms on the Sun:
http://www.swpc.noaa.gov/sites/default/files/images/u33/swx_booklet.pdf. An illustrated booklet from NOAA.

Stanford Solar Center: <http://solar-center.stanford.edu/>. An excellent site with information for students and teachers.

Apps

These can tell you and your students more about what's happening on the Sun in real time.

NASA Space Weather: <https://itunes.apple.com/us/app/nasa-space-weather/id422621403?mt=8>.

Solaris Alpha: <https://play.google.com/store/apps/details?id=com.tomoreilly.solarisalpha>.

Videos

Journey into the Sun: <https://www.youtube.com/watch?v=fqKFQ7z0Nuk>. 2010 KQED Quest TV Program mostly about the Solar Dynamics Observatory spacecraft, its launch and capabilities, but with good general information on how the Sun works (12:24).

NASA | SDO: Three Years in Three Minutes--With Expert Commentary: <https://www.youtube.com/watch?v=QaCG0wAjJSY&src>. Video of 3 years of observations of the Sun by the Solar Dynamics Observatory made into a speeded up movie, with commentary by solar physicist Alex Young (5:03).

Our Explosive Sun: <http://www.youtube.com/watch?v=kI6YGSIJqrE>. Video of a 2011 public lecture in the Silicon Valley Astronomy Lecture Series by Dr. Thomas Berger about solar activity and recent satellite missions to observe and understand it (1:20:22).

Out There Raining Fire:
<http://www.nytimes.com/video/science/100000003489464/out-there-raining-fire.html?>

[emc=eta1](#). Nice overview and introduction to the Sun by science reporter Dennis Overbye of the NY Times (2:28)

Space Weather Impacts: <http://www.swpc.noaa.gov/content/education-and-outreach>. Video from NOAA (2:47); <https://www.youtube.com/playlist?list=PLBdd8cMH5jFmvVR2sZubIUzBO6JI0Pvx0>. Videos from the National Weather Service (four short videos) (14:41).

Space Weather: Storms on the Sun: <http://www.youtube.com/watch?v=vWsmp4o-qVg>. Science bulletin from the American Museum of Natural History, giving the background to what happens on the Sun to cause space weather (6:10).

Sun Storms: <http://www.livescience.com/11754-sun-storms-havoc-electronic-world.html>. From the Starry Night company about storms from the Sun now and in the past (4:49).

Sunspot Group AR 2339 Crosses the Sun: <http://apod.nasa.gov/apod/ap150629.html>. Short video (with music) animates Solar Dynamics Observatory images of an especially large sunspot group going across the Sun's face (1:15).

What Happens on the Sun Doesn't Stay on the Sun: https://www.youtube.com/watch?v=bg_gD2-ujCk. From the National Oceanic and Atmospheric Administration: introduction to the Sun, space weather, its effects, and how we monitor it (4:56).

Collaborative Group Activities

- A. Have your group make a list of all the ways the Sun personally affects your life on Earth. (Consider the everyday effects as well as the unusual effects due to high solar activity.)
- B. Long before the nature of the Sun was fully understood, astronomer (and planet discoverer) William Herschel (1738–1822) proposed that the hot Sun may have a cool interior and may be inhabited. Have your group discuss this proposal and come up with modern arguments against it.
- C. We discussed how the migration of Europeans to North America was apparently affected by short-term climate change. If Earth were to become significantly hotter, either because of changes in the Sun or because of greenhouse warming, one effect would be an increase in the rate of melting of the polar ice caps. How would this affect modern civilization?
- D. Suppose we experience another Maunder Minimum on Earth, and it is accompanied by a drop in the average temperature like the Little Ice Age in Europe. Have your group discuss how this would affect civilization and

international politics. Make a list of the most serious effects that you can think of.

- E. Watching sunspots move across the disk of the Sun is one way to show that our star rotates on its axis. Can your group come up with other ways to show the Sun's rotation?
- F. Suppose in the future, we are able to forecast space weather as well as we forecast weather on Earth. And suppose we have a few days of warning that a big solar storm is coming that will overload Earth's magnetosphere with charged particles and send more ultraviolet and X-rays toward our planet. Have your group discuss what steps we might take to protect our civilization?
- G. Have your group members research online to find out what satellites are in space to help astronomers study the Sun. In addition to searching for NASA satellites, you might also check for satellites launched by the European Space Agency and the Japanese Space Agency.
- H. Some scientists and engineers are thinking about building a "solar sail"—something that can use the Sun's wind or energy to propel a spacecraft away from the Sun. The Planetary Society is a nonprofit organization that is trying to get solar sails launched, for example. Have your group do a report on the current state of solar-sail projects and what people are dreaming about for the future.

Review Questions

Exercise:

Problem:

Describe the main differences between the composition of Earth and that of the Sun.

Exercise:

Problem:

Describe how energy makes its way from the nuclear core of the Sun to the atmosphere. Include the name of each layer and how energy moves through the layer.

Exercise:

Problem:

Make a sketch of the Sun's atmosphere showing the locations of the photosphere, chromosphere, and corona. What is the approximate temperature of each of these regions?

Exercise:

Problem: Why do sunspots look dark?

Exercise:

Problem:

Which aspects of the Sun's activity cycle have a period of about 11 years? Which vary during intervals of about 22 years?

Exercise:

Problem:

Summarize the evidence indicating that over several hundreds of years or more there have been variations in the level of the solar activity.

Exercise:

Problem: What is the Zeeman effect and what does it tell us about the Sun?

Exercise:

Problem:

Explain how the theory of the Sun's dynamo results in an average 22-year solar activity cycle. Include the location and mechanism for the dynamo.

Exercise:

Problem:

Compare and contrast the four different types of solar activity above the photosphere.

Exercise:

Problem:

What are the two sources of particles coming from the Sun that cause space weather? How are they different?

Exercise:

Problem:

How does activity on the Sun affect human technology on Earth and in the rest of the solar system?

Exercise:

Problem: How does activity on the Sun affect natural phenomena on Earth?

Thought Questions

Exercise:

Problem:

[\[link\]](#) indicates that the density of the Sun is 1.41 g/cm^3 . Since other materials, such as ice, have similar densities, how do you know that the Sun is not made of ice?

Exercise:

Problem:

Starting from the core of the Sun and going outward, the temperature decreases. Yet, above the photosphere, the temperature increases. How can this be?

Exercise:

Problem:

Since the rotation period of the Sun can be determined by observing the apparent motions of sunspots, a correction must be made for the orbital motion of Earth. Explain what the correction is and how it arises. Making some sketches may help answer this question.

Exercise:

Problem:

Suppose an (extremely hypothetical) elongated sunspot forms that extends from a latitude of 30° to a latitude of 40° along a fixed longitude on the Sun. How will the appearance of that sunspot change as the Sun rotates? ([\[link\]](#) should help you figure this out.)

Exercise:

Problem:

The text explains that plagues are found near sunspots, but [\[link\]](#) shows that they appear even in areas without sunspots. What might be the explanation for this?

Exercise:

Problem:

Why would a flare be observed in visible light, when they are so much brighter in X-ray and ultraviolet light?

Exercise:**Problem:**

How can the prominences, which are so big and ‘float’ in the corona, stay gravitationally attached to the Sun while flares can escape?

Exercise:**Problem:**

If you were concerned about space weather and wanted to avoid it, where would be the safest place on Earth for you to live?

Exercise:**Problem:**

Suppose you live in northern Canada and an extremely strong flare is reported on the Sun. What precautions might you take? What might be a positive result?

Figuring for Yourself**Exercise:****Problem:**

The edge of the Sun doesn’t have to be absolutely sharp in order to look that way to us. It just has to go from being transparent to being completely opaque in a distance that is smaller than your eye can resolve. Remember from [Astronomical Instruments](#) that the ability to resolve detail depends on the size of the telescope’s aperture. The pupil of your eye is very small relative to the size of a telescope and therefore is very limited in the amount of detail you can see. In fact, your eye cannot see details that are smaller than $1/30$ of the diameter of the Sun (about 1 arcminute). Nearly all the light from the Sun emerges from a layer that is only about 400 km thick. What fraction is this of the diameter of the Sun? How does this compare with the ability of the human eye to resolve detail? Suppose we could see light emerging directly from a layer that was 300,000 km thick. Would the Sun appear to have a sharp edge?

Exercise:

Problem:

Show that the statement that 92% of the Sun's atoms are hydrogen is consistent with the statement that 73% of the Sun's mass is made up of hydrogen, as found in [\[link\]](#). (Hint: Make the simplifying assumption, which is nearly correct, that the Sun is made up entirely of hydrogen and helium.)

Exercise:**Problem:**

From Doppler shifts of the spectral lines in the light coming from the east and west edges of the Sun, astronomers find that the radial velocities of the two edges differ by about 4 km/s, meaning that the Sun's rotation rate is 2 km/s. Find the approximate period of rotation of the Sun in days. The circumference of a sphere is given by $2\pi R$, where R is the radius of the sphere.

Exercise:**Problem:**

Assuming an average sunspot cycle of 11 years, how many revolutions does the equator of the Sun make during that one cycle? Do higher latitudes make more or fewer revolutions compared to the equator?

Exercise:**Problem:**

This chapter gives the average sunspot cycle as 11 years. Verify this using [\[link\]](#).

Exercise:**Problem:**

The escape velocity from any astronomical object can be calculated as $v_{\text{escape}} = \sqrt{2GM/R}$. Using the data in [Appendix E](#), calculate the escape velocity from the photosphere of the Sun. Since coronal mass ejections escape from the corona, would the escape velocity from there be more or less than from the photosphere?

Exercise:**Problem:**

Suppose you observe a major solar flare while astronauts are orbiting Earth. Use the data in the text to calculate how long it will before the charged particles ejected from the Sun during the flare reach them.

Exercise:**Problem:**

Suppose an eruptive prominence rises at a speed of 150 km/s. If it does not change speed, how far from the photosphere will it extend after 3 hours? How does this distance compare with the diameter of Earth?

Exercise:**Problem:**

From the information in [\[link\]](#), estimate the speed with which the particles in the CME in parts (c) and (d) are moving away from the Sun.

Sources of Sunshine: Thermal and Gravitational Energy

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Identify different forms of energy
- Understand the law of conservation of energy
- Explain ways that energy can be transformed

Energy is a challenging concept to grasp because it exists in so many different forms that it defies any single simple explanation. In many ways, comprehending energy is like comprehending wealth: There are very different forms of wealth and they follow different rules, depending on if they are the stock market, real estate, a collection of old comic books, great piles of cash, or one of the many other ways to make and lose money. It is easier to discuss one or two forms of wealth—or energy—than to discuss that concept in general.

When striving to understand how the Sun can put out so much energy for so long, scientists considered many different types of energy. Nineteenth-century scientists knew of two possible sources for the Sun's energy: chemical and gravitational energy. The source of chemical energy most familiar to them was the burning (the chemical term is *oxidation*) of wood, coal, gasoline, or other fuel. We know exactly how much energy the burning of these materials can produce. We can thus calculate that even if the immense mass of the Sun consisted of a burnable material like coal or wood, our star could not produce energy at its present rate for more than few thousand years. However, we know from geologic evidence that water

was present on Earth's surface nearly 4 billion years ago, so the Sun must have been shining brightly (and making Earth warm) at least as long as that. Today, we also know that at the temperatures found in the Sun, nothing like solid wood or coal could survive.

Note:

What's Watt?

Just a word about the units we are using. A watt (W) is a unit of *power*, which is energy used or given off per unit time. It is measured in joules per second (J/s). You know from your everyday experience that it is not just *how much* energy you expend, but *how long* you take to do it. (Burning 10 Calories in 10 minutes requires a very different kind of exercise than burning those 10 Calories in an hour.) Watts tell you the *rate* at which energy is being used; for example, a 100-watt bulb uses 100 joules (J) of energy every second.

And how big is a joule? A 73-kilogram (160-pound) astronomy instructor running at about 4.4 meters per second (10 miles per hour) because he is late for class has a motion energy of about 700 joules.

Conservation of Energy

Other nineteenth-century attempts to determine what makes the Sun shine used the law of conservation of energy. Simply stated, this law says that energy cannot be created or destroyed, but can be transformed from one type to another, such as from heat to mechanical energy. The steam engine, which was key to the Industrial Revolution, provides a good example. In this type of engine, the hot steam from a boiler drives the movement of a piston, converting heat energy into motion energy.

Conversely, motion can be transformed into heat. If you clap your hands vigorously at the end of an especially good astronomy lecture, your palms become hotter. If you rub ice on the surface of a table, the heat produced by friction melts the ice. The brakes on cars use friction to reduce speed, and in the process, transform motion energy into heat energy. That is why after

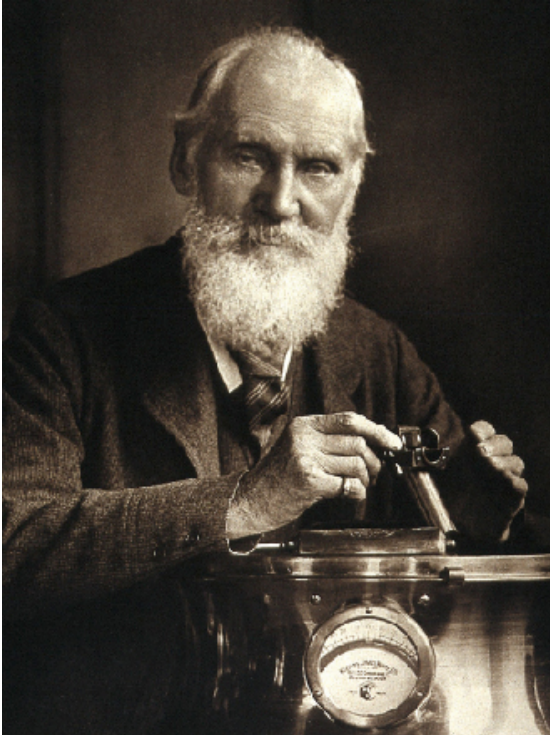
bringing a car to a stop, the brakes can be very hot; this also explains why brakes can overheat when used carelessly while descending long mountain roads.

In the nineteenth century, scientists thought that the source of the Sun's heat might be the mechanical motion of meteorites falling into it. Their calculations showed, however, that in order to produce the total amount of energy emitted by the Sun, the mass in meteorites that would have to fall into the Sun every 100 years would equal the mass of Earth. The resulting increase in the Sun's mass would, according to Kepler's third law, change the period of Earth's orbit by 2 seconds per year. Such a change would be easily measurable and was not, in fact, occurring. Scientists could then disprove this as the source of the Sun's energy.

Gravitational Contraction as a Source of Energy

Proposing an alternative explanation, British physicist Lord Kelvin and German scientist Hermann von Helmholtz ([link](#)), in about the middle of the nineteenth century, proposed that the Sun might produce energy by the conversion of gravitational energy into heat. They suggested that the outer layers of the Sun might be “falling” inward because of the force of gravity. In other words, they proposed that the Sun could be shrinking in size, staying hot and bright as a result.

Kelvin (1824–1907) and Helmholtz (1821–1894).



(a)



(b)

(a) British physicist William Thomson (Lord Kelvin) and (b) German scientist Hermann von Helmholtz proposed that the contraction of the Sun under its own gravity might account for its energy. (credit a: modification of work by Wellcome Library, London; credit b: modification of work by Wellcome Library, London)

To imagine what would happen if this hypothesis were true, picture the outer layer of the Sun starting to fall inward. This outer layer is a gas made up of individual atoms, all moving about in random directions. If a layer falls inward, the atoms acquire an additional speed because of falling motion. As the outer layer falls inward, it also contracts, moving the atoms closer together. Collisions become more likely, and some of them transfer the extra speed associated with the falling motion to other atoms. This, in turn, increases the speeds of those atoms. The temperature of a gas is a measure of the kinetic energy (motion) of the atoms within it; hence, the temperature of this layer of the Sun increases. Collisions also excite electrons within the atoms to higher-energy orbits. When these electrons

return to their normal orbits, they emit photons, which can then escape from the Sun (see [Radiation and Spectra](#)).

Kelvin and Helmholtz calculated that a contraction of the Sun at a rate of only about 40 meters per year would be enough to produce the amount of energy that it is now radiating. Over the span of human history, the decrease in the Sun's size from such a slow contraction would be undetectable.

If we assume that the Sun began its life as a large, diffuse cloud of gas, then we can calculate how much energy has been radiated by the Sun during its entire lifetime as it has contracted from a very large diameter to its present size. The amount of energy is on the order of 10^{42} joules. Since the solar luminosity is 4×10^{26} watts (joules/second) or about 10^{34} joules per year, contraction could keep the Sun shining at its present rate for roughly 100 million years.

In the nineteenth century, 100 million years at first seemed plenty long enough, since Earth was then widely thought to be much younger than this. But toward the end of that century and into the twentieth, geologists and physicists showed that Earth (and, hence, the Sun) is actually much older. Contraction therefore cannot be the primary source of solar energy (although, as we shall see in [The Birth of Stars and the Discovery of Planets Outside the Solar System](#), contraction is an important source of energy for a while in stars that are just being born). Scientists were thus confronted with a puzzle of enormous proportions. Either an unknown type of energy was responsible for the most important energy source known to humanity, or estimates of the age of the solar system (and life on Earth) had to be seriously modified. Charles Darwin, whose theory of evolution required a longer time span than the theories of the Sun seemed to permit, was discouraged by these results and continued to worry about them until his death in 1882.

It was only in the twentieth century that the true source of the Sun's energy was identified. The two key pieces of information required to solve the puzzle were the structure of the nucleus of the atom and the fact that mass can be converted into energy.

Key Concepts and Summary

The Sun produces an enormous amount of energy every second. Since Earth and the solar system are roughly 4.5 billion years old, this means that the Sun has been producing vast amounts of energy for a very, very long time. Neither chemical burning nor gravitational contraction can account for the total amount of energy radiated by the Sun during all this time.

Mass, Energy, and the Theory of Relativity

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how matter can be converted into energy
- Describe the particles that make up atoms
- Describe the nucleus of an atom
- Understand the nuclear forces that hold atoms together
- Trace the nuclear reactions in the solar interior

As we have seen, energy cannot be created or destroyed, but only converted from one form to another. One of the remarkable conclusions derived by Albert Einstein (see [Albert Einstein](#)) when he developed his theory of relativity is that matter can be considered a form of energy too and can be converted into energy. Furthermore, energy can also be converted into matter. This seemed to contradict what humans had learned over thousands of years by studying nature. Matter is something we can see and touch, whereas energy is something objects have when they do things like move or heat up. The idea that matter or energy can be converted into each other seemed as outrageous as saying you could accelerate a car by turning the bumper into more speed, or that you could create a bigger front seat by slowing down your car. That would be pretty difficult to believe; yet, the universe actually works somewhat like that.

Converting Matter into Energy

The remarkable equivalence between matter and energy is given in one of the most famous equations:

Equation:

$$E = mc^2$$

In this equation, E stands for energy, m stands for mass, and c , the constant that relates the two, is the speed of light (3×10^8 meters per second). Note that mass is a measure of the quantity of matter, so the significance of this equation is that matter can be converted into energy and energy can be converted into matter. Let's compare this equation of converting matter and energy to some common conversion equations that have the same form:

Equation:

$$\text{inches} = \text{feet} \times 12, \text{ or cents} = \text{dollars} \times 100$$

Just as each conversion formula allows you to calculate the conversion of one thing into another, when we convert matter into energy, we consider how much mass the matter has. The conversion factor in this case turns out not to be either 12 or 100, as in our examples, but another constant quantity: the speed of light squared. Note that matter does not have to travel at the speed of light (or the speed of light squared) for this conversion to occur. The factor of c^2 is just the number that Einstein showed must be used to relate mass and energy.

Notice that this formula does not tell us *how* to convert mass into energy, just as the formula for cents does not tell us where to exchange coins for a dollar bill. The formulas merely tell us what the equivalent values are if we succeed in making the conversion. When Einstein first derived his formula in 1905, no one had the faintest idea how to convert mass into energy in any practical way. Einstein himself tried to discourage speculation that the large-scale conversion of atomic mass into energy would be feasible in the near future. Today, as a result of developments in nuclear physics, we regularly convert mass into energy in power plants, nuclear weapons, and high-energy physics experiments in particle accelerators.

Because the speed of light squared (c^2) is a very large quantity, the conversion of even a small amount of mass results in a very large amount of energy. For example, the complete conversion of 1 gram of matter (about 1/28 ounce, or approximately 1 paperclip) would produce as much energy as the burning of 15,000 barrels of oil.

Scientists soon realized that the conversion of mass into energy is the source of the Sun's heat and light. With Einstein's equation of $E = mc^2$, we can calculate that the amount of energy radiated by the Sun could be produced by the complete conversion of about 4 million tons of matter into energy inside the Sun each second. Destroying 4 million tons per second sounds like a lot when compared to earthly things, but bear in mind that the Sun is a very big reservoir of matter. In fact, we will see that the Sun contains more than enough mass to destroy such huge amounts of matter and still continue shining at its present rate for billions of years.

But knowing all that still does not tell us *how* mass can be converted into energy. To understand the process that actually occurs in the Sun, we need to explore the structure of the atom a bit further.

Note:**Albert Einstein**

For a large part of his life, Albert Einstein ([link](#)) was one of the most recognized celebrities of his day. Strangers stopped him on the street, and people all over the world asked him for endorsements, advice, and assistance. In fact, when Einstein and the great film star Charlie Chaplin met in California, they found they shared similar feelings about the loss of privacy that came with fame. Einstein's name was a household word despite the fact that most people did not understand the ideas that had made him famous.

Einstein was born in 1879 in Ulm, Germany. Legend has it that he did not do well in school (even in arithmetic), and thousands of students have since attempted to justify a bad grade by referring to this story. Alas, like many legends, this one is not true. Records indicate that although he tended to rebel against the authoritarian teaching style in vogue in Germany at that time, Einstein was a good student.

After graduating from the Federal Polytechnic Institute in Zurich, Switzerland, Einstein at first had trouble getting a job (even as a high school teacher), but he eventually became an examiner in the Swiss Patent Office. Working in his spare time, without the benefit of a university environment but using his superb physical intuition, he wrote four papers in 1905 that would ultimately transform the way physicists looked at the world.

One of these, which earned Einstein the Nobel Prize in 1921, set part of the foundation of *quantum mechanics*—the rich, puzzling, and remarkable theory of the subatomic realm. But his most important paper presented the *special theory of relativity*, a reexamination of space, time, and motion that added a whole new level of sophistication to our understanding of those concepts. The famed equation $E = mc^2$ was actually a relatively minor part of this theory, added in a later paper.

In 1916, Einstein published his *general theory of relativity*, which was, among other things, a fundamentally new description of gravity (see [Black Holes and Curved Spacetime](#)). When this theory was confirmed by measurements of the “bending of starlight” during a 1919 eclipse (*The New York Times* headline read, “Lights All Askew in the Heavens”), Einstein became world famous.

In 1933, to escape Nazi persecution, Einstein left his professorship in Berlin and settled in the United States at the newly created Institute for Advanced Studies at Princeton. He remained there until his death in 1955, writing, lecturing, and espousing a variety of intellectual and political causes. For example, he agreed to sign a letter written by Leo Szilard and other scientists in 1939, alerting President Roosevelt to the dangers of allowing Nazi Germany to develop the atomic bomb first. And in 1952, Einstein was offered the second presidency of Israel. In declining the position, he said, “I know a little about nature and hardly anything about men.”

Albert Einstein (1879–1955).



This portrait of Einstein was taken in 1912. (credit: modification of work by J. F. Langhans)

Elementary Particles

The fundamental components of atoms are the proton, neutron, and electron (see [The Structure of the Atom](#)).

Protons, neutrons, and electrons are by no means all the particles that exist. First, for each kind of particle, there is a corresponding but opposite *antiparticle*. If the particle carries a charge, its antiparticle has the opposite charge. The antielectron is the *positron*, which has the same mass as the electron but is positively charged. Similarly, the antiproton has a negative charge. The remarkable thing about such *antimatter* is that when a particle

comes into contact with its antiparticle, the original particles are annihilated, and substantial amounts of energy in the form of photons are produced.

Since our world is made exclusively of ordinary particles of matter, antimatter cannot survive for very long. But individual antiparticles are found in cosmic rays (particles that arrive at the top of Earth's atmosphere from space) and can be created in particle accelerators. And, as we will see in a moment, antimatter is created in the core of the Sun and other stars.

Science fiction fans may be familiar with antimatter from the *Star Trek* television series and films. The Starship Enterprise is propelled by the careful combining of matter and antimatter in the ship's engine room. According to $E = mc^2$, the annihilation of matter and antimatter can produce a huge amount of energy, but keeping the antimatter fuel from touching the ship before it is needed must be a big problem. No wonder Scotty, the chief engineer in the original TV show, always looked worried!

In 1933, physicist Wolfgang Pauli ([\[link\]](#)) suggested that there might be another type of elementary particle. Energy seemed to disappear when certain types of nuclear reactions took place, violating the law of conservation of energy. Pauli was reluctant to accept the idea that one of the basic laws of physics was wrong, and he suggested a “desperate remedy.” Perhaps a so-far-undetected particle, which was given the name **neutrino** (“little neutral one”), carried away the “missing” energy. He suggested that neutrinos were particles with zero mass, and that like photons, they moved with the speed of light.

Wolfgang Pauli in 1945.



Pauli is considered the “father” of the neutrino, having conceived of it in 1933.

The elusive neutrino was not detected until 1956. The reason it was so hard to find is that neutrinos interact very weakly with other matter and therefore are very difficult to detect. Earth is more transparent to a neutrino than the thinnest and cleanest pane of glass is to a photon of light. In fact, most neutrinos can pass completely through a star or planet without being absorbed. As we shall see, this behavior of neutrinos makes them a very important tool for studying the Sun. Since Pauli’s prediction, scientists have learned a lot more about the neutrino. We now know that there are three different types of neutrinos, and in 1998, neutrinos were discovered to have a tiny amount of mass. Indeed, it is so tiny that electrons are at least

500,000 times more massive. Ongoing research is focused on determining the mass of neutrinos more precisely, and it may still turn out that one of the three types is massless. We will return to the subject of neutrinos later in this chapter.

Some of the properties of the proton, electron, neutron, and neutrino are summarized in [\[link\]](#). (Other subatomic particles have been produced by experiments with particle accelerators, but they do not play a role in the generation of solar energy.)

Properties of Some Common Particles		
Particle	Mass (kg)	Charge
Proton	1.67265×10^{-27}	+1
Neutron	1.67495×10^{-27}	0
Electron	9.11×10^{-31}	-1
Neutrino	$<2 \times 10^{-36}$ (uncertain)	0

The Atomic Nucleus

The nucleus of an atom is not just a loose collection of elementary particles. Inside the nucleus, particles are held together by a very powerful force called the strong nuclear force. This is short-range force, only capable of acting over distances about the size of the atomic nucleus. A quick thought experiment shows how important this force is. Take a look at your finger and consider the atoms composing it. Among them is carbon, one of the basic elements of life. Focus your imagination on the nucleus of one of your carbon atoms. It contains six protons, which have a positive charge, and six

neutrons, which are neutral. Thus, the nucleus has a net charge of six positives. If only the electrical force were acting, the protons in this and every carbon atom would find each other very repulsive and fly apart.

The strong nuclear force is an attractive force, stronger than the electrical force, and it keeps the particles of the nucleus tightly bound together. We saw earlier that if under the force of gravity a star “shrinks”—bringing its atoms closer together—gravitational energy is released. In the same way, if particles come together under the strong nuclear force and unite to form an atomic nucleus, some of the nuclear energy is released. The energy given up in such a process is called the *binding energy* of the nucleus.

When such binding energy is released, the resulting nucleus has slightly less mass than the sum of the masses of the particles that came together to form it. In other words, the energy comes from the loss of mass. This slight deficit in mass is only a small fraction of the mass of one proton. But because each bit of lost mass can provide a lot of energy (remember, $E = mc^2$), this nuclear energy release can be quite substantial.

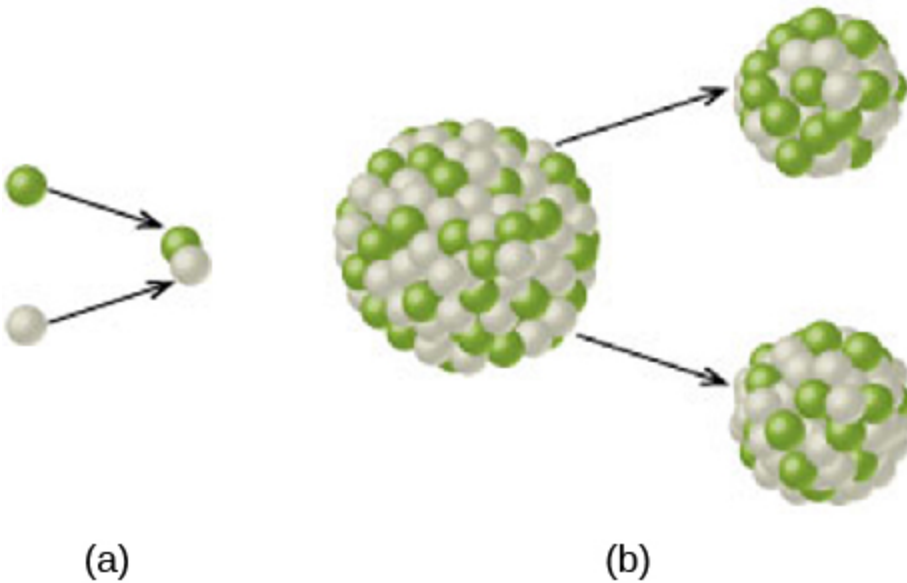
Measurements show that the binding energy is greatest for atoms with a mass near that of the iron nucleus (with a combined number of protons and neutrons equal to 56) and less for both the lighter and the heavier nuclei. Iron, therefore, is the most stable element: since it gives up the most energy when it forms, it would require the most energy to break it back down into its component particles.

What this means is that, in general, when light atomic nuclei come together to form a heavier one (up to iron), mass is lost and energy is released. This joining together of atomic nuclei is called nuclear **fusion**.

Energy can also be produced by breaking up heavy atomic nuclei into lighter ones (down to iron); this process is called nuclear **fission**. Nuclear fission was the process we learned to use first—in atomic bombs and in nuclear reactors used to generate electrical power—and it may therefore be more familiar to you. Fission also sometimes occurs spontaneously in some unstable nuclei through the process of natural radioactivity. But fission requires big, complex nuclei, whereas we know that the stars are made up

predominantly of small, simple nuclei. So we must look to fusion first to explain the energy of the Sun and the stars ([link](#)).

Fusion and Fission.



(a) In fusion, light atomic nuclei join together to form a heavier nuclei, releasing energy in the process. (b) In fission, energy is produced by the breaking up of heavy, complex nuclei into lighter ones.

Nuclear Attraction versus Electrical Repulsion

So far, we seem to have a very attractive prescription for producing the energy emitted by the Sun: “roll” some nuclei together and join them via nuclear fusion. This will cause them to lose some of their mass, which then turns into energy. However, every nucleus, even simple hydrogen, has protons—and protons all have positive charges. Since like charges repel via the electrical force, the closer we get two nuclei to each other, the more they repel. It’s true that if we can get them within “striking distance” of the nuclear force, they will then come together with a much stronger attraction. But that striking distance is very tiny, about the size of a nucleus. How can we get nuclei close enough to participate in fusion?

The answer turns out to be heat—tremendous heat—which speeds the protons up enough to overcome the electrical forces that try to keep protons apart. Inside the Sun, as we saw, the most common element is hydrogen, whose nucleus contains only a single proton. Two protons can fuse only in regions where the temperature is greater than about 12 million K, and the speed of the protons average around 1000 kilometers per second or more. (In old-fashioned units, that’s over 2 million miles per hour!)

In our Sun, such extreme temperatures are reached only in the regions near its center, which has a temperature of 15 million K. Calculations show that nearly all of the Sun’s energy is generated within about 150,000 kilometers of its core, or within less than 10% of its total volume.

Even at these high temperatures, it is exceedingly difficult to force two protons to combine. On average, a proton will rebound from other protons in the Sun’s crowded core for about 14 billion years, at the rate of *100 million collisions per second*, before it fuses with a second proton. This is, however, only the *average* waiting time. Some of the enormous numbers of protons in the Sun’s inner region are “lucky” and take only a few collisions to achieve a fusion reaction: they are the protons responsible for producing the energy radiated by the Sun. Since the Sun is about 4.5 billion years old, most of its protons have not yet been involved in fusion reactions.

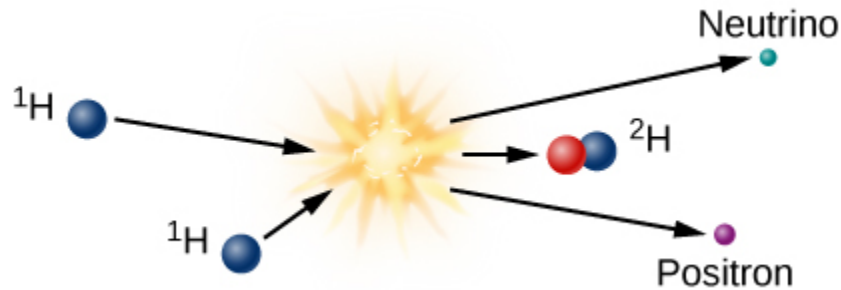
Nuclear Reactions in the Sun’s Interior

The Sun, then, taps the energy contained in the nuclei of atoms through nuclear fusion. Let’s look at what happens in more detail. Deep inside the Sun, a three-step process takes four hydrogen nuclei and fuses them together to form a single helium nucleus. The helium nucleus is slightly less massive than the four hydrogen nuclei that combine to form it, and that mass is converted into energy.

The initial step required to form one helium nucleus from four hydrogen nuclei is shown in [\[link\]](#). At the high temperatures inside the Sun’s core, two protons combine to make a *deuterium* nucleus, which is an isotope (or version) of hydrogen that contains one proton and one neutron. In effect, one of the original protons has been converted into a neutron in the fusion

reaction. Electric charge has to be conserved in nuclear reactions, and it is conserved in this one. A **positron** (antimatter electron) emerges from the reaction and carries away the positive charge originally associated with one of the protons.

Proton-Proton Chain, Step 1.



This is the first step in the process of fusing hydrogen into helium in the Sun. High temperatures are required because this reaction starts with two hydrogen nuclei, which are protons (shown in blue at left) that must overcome electrical repulsion to combine, forming a hydrogen nucleus with a proton and a neutron (shown in red). Note that hydrogen containing one proton and one neutron is given its own name: deuterium. Also produced in this reaction are a positron, which is an antielectron, and an elusive particle named the neutrino.

Since it is antimatter, this positron will instantly collide with a nearby electron, and both will be annihilated, producing electromagnetic energy in the form of gamma-ray photons. This gamma ray, which has been created in the center of the Sun, finds itself in a world crammed full of fast-moving nuclei and electrons. The gamma ray collides with particles of matter and transfers its energy to one of them. The particle later emits another gamma-ray photon, but often the emitted photon has a bit less energy than the one that was absorbed.

Such interactions happen to gamma rays again and again and again as they make their way slowly toward the outer layers of the Sun, until their energy becomes so reduced that they are no longer gamma rays but X-rays (recall what you learned in [The Electromagnetic Spectrum](#)). Later, as the photons lose still more energy through collisions in the crowded center of the Sun, they become ultraviolet photons.

By the time they reach the Sun's surface, most of the photons have given up enough energy to be ordinary light—and they are the sunlight we see coming from our star. (To be precise, each gamma-ray photon is ultimately converted into many separate lower-energy photons of sunlight.) So, the sunlight given off by the Sun today had its origin as a gamma ray produced by nuclear reactions deep in the Sun's core. The length of time that photons require to reach the surface depends on how far a photon on average travels between collisions, and the travel time depends on what model of the complicated solar interior we accept. Estimates are somewhat uncertain but indicate that the emission of energy from the surface of the Sun can lag its production in the interior by 100,000 years to as much as 1,000,000 years.

In addition to the positron, the fusion of two hydrogen atoms to form deuterium results in the emission of a neutrino. Because neutrinos interact so little with ordinary matter, those produced by fusion reactions near the center of the Sun travel directly to the Sun's surface and then out into space, in all directions. Neutrinos move at nearly the speed of light, and they escape the Sun about two seconds after they are created.

Proton-Proton Chain, Step 2.



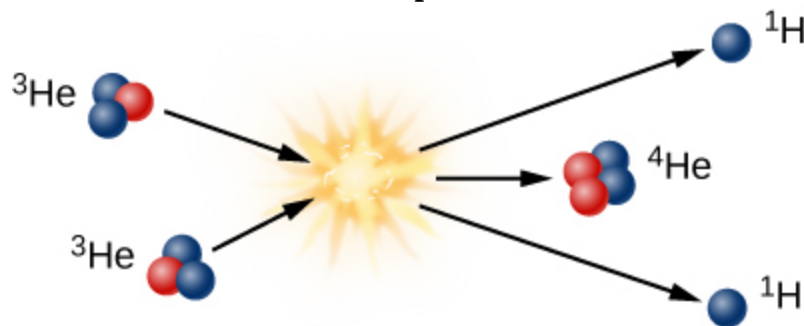
This is the second step of the proton-proton chain, the fusion reaction that converts hydrogen into helium in the Sun.

This step combines one hydrogen nucleus, which is a proton (shown in blue), with the deuterium nucleus from the previous step (shown as a red and blue particle). The product of this is an isotope of helium with two protons (blue) and one neutron (red) and energy in the form of gamma-ray radiation.

The second step in forming helium from hydrogen is to add another proton to the deuterium nucleus to create a helium nucleus that contains two protons and one neutron ([\[link\]](#)). In the process, some mass is again lost and more gamma radiation is emitted. Such a nucleus is helium because an element is defined by its number of protons; any nucleus with two protons is called helium. But this form of helium, which we call helium-3 (and write in shorthand as ^3He) is not the isotope we see in the Sun's atmosphere or on Earth. That helium has two neutrons and two protons and hence is called helium-4 (^4He).

To get to helium-4 in the Sun, helium-3 must combine with another helium-3 in the third step of fusion (illustrated in [\[link\]](#)). Note that two energetic protons are left over from this step; each of them comes out of the reaction ready to collide with other protons and to start step 1 in the chain of reactions all over again.

Proton-Proton Chain, Step 3.



This is the third step in the fusion of hydrogen into helium in the Sun. Note that

the two helium-3 nuclei from the second step (see [\[link\]](#)) must combine before the third step becomes possible. The two protons that come out of this step have the energy to collide with other protons in the Sun and start step one again.

Note:

These animations of [proton-proton reactions](#) show the steps required for fusion of hydrogen into helium in the Sun.

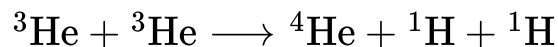
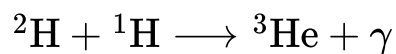
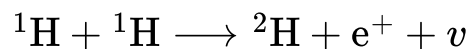
Note:

Visit the [Tokamak Fusion Reactor](#) at the General Atomics Lab in San Diego, CA, for an 8-minute tour.

The Proton-Proton Chain

The nuclear reactions in the Sun that we have been discussing can be described succinctly through the following nuclear formulas:

Equation:



Here, the superscripts indicate the total number of neutrons plus protons in the nucleus, e^+ is the positron, ν is the neutrino, and γ indicates that gamma

rays are emitted. Note that the third step requires two helium-3 nuclei to start; the first two steps must happen twice before the third step can occur.

Although, as we discussed, the first step in this chain of reactions is very difficult and generally takes a long time, the other steps happen more quickly. After the deuterium nucleus is formed, it survives an average of only about 6 seconds before being converted into ^3He . About a million years after that (on average), the ^3He nucleus will combine with another to form ^4He .

We can compute the amount of energy these reactions generate by calculating the difference in the initial and final masses. The masses of hydrogen and helium atoms in the units normally used by scientists are $1.007825\ u$ and $4.00268\ u$, respectively. (The unit of mass, u , is defined to be $1/12$ the mass of an atom of carbon, or approximately the mass of a proton.) Here, we include the mass of the entire atom, not just the nucleus, because electrons are involved as well. When hydrogen is converted into helium, two positrons are created (remember, the first step happens twice), and these are annihilated with two free electrons, adding to the energy produced.

Equation:

$$\begin{aligned} 4 \times 1.007825 &= 4.03130\ u \text{ (mass of initial hydrogen atoms)} \\ &- 4.00268\ u \text{ (mass of final helium atoms)} \\ &= 0.02862\ u \text{ (mass lost in the transformation)} \end{aligned}$$

The mass lost, $0.02862\ u$, is 0.71% of the mass of the initial hydrogen. Thus, if 1 kilogram of hydrogen is converted into helium, then the mass of the helium is only 0.9929 kilogram, and 0.0071 kilogram of material is converted into energy. The speed of light (c) is 3×10^8 meters per second, so the energy released by the conversion of just 1 kilogram of hydrogen into helium is:

Equation:

$$E = mc^2$$

$$E = 0.0071 \text{ kg} \times 3 \times 10^8 \text{ m/s}^2 = 6.4 \times 10^{14} \text{ J}$$

This amount, the energy released when a single kilogram (2.2 pounds) of hydrogen undergoes fusion, would supply all the electricity used by the typical U.S. household for roughly 17,000 years.

To produce the Sun's luminosity of 4×10^{26} watts, some 600 million tons of hydrogen must be converted into helium *each second*, of which about 4 million tons are converted from matter into energy. As large as these numbers are, the store of hydrogen (and thus of nuclear energy) in the Sun is still *more* enormous, and can last a long time—billions of years, in fact.

At the temperatures inside the stars with masses smaller than about 1.2 times the mass of our Sun (a category that includes the Sun itself), most of the energy is produced by the reactions we have just described, and this set of reactions is called the **proton-proton chain** (or sometimes, the p-p chain). In the proton-proton chain, protons collide directly with other protons to form helium nuclei.

In hotter stars, another set of reactions, called the carbon-nitrogen-oxygen (CNO) cycle, accomplishes the same net result. In the CNO cycle, carbon and hydrogen nuclei collide to initiate a series of reactions that form nitrogen, oxygen, and ultimately, helium. The nitrogen and oxygen nuclei do not survive but interact to form carbon again. Therefore, the outcome is the same as in the proton-proton chain: four hydrogen atoms disappear, and in their place, a single helium atom is created. The CNO cycle plays only a minor role in the Sun but is the main source of energy for stars with masses greater than about the mass of the Sun.

So you can see that we have solved the puzzle that so worried scientists at the end of the nineteenth century. The Sun can maintain its high temperature and energy output for billions of years through the fusion of the simplest element in the universe, hydrogen. Because most of the Sun (and the other stars) is made of hydrogen, it is an ideal “fuel” for powering a star. As will be discussed in the following chapters, we can define a star as a ball

of gas capable of getting its core hot enough to initiate the fusion of hydrogen. There are balls of gas that lack the mass required to do this (Jupiter is a local example); like so many hopefuls in Hollywood, they will never be stars.

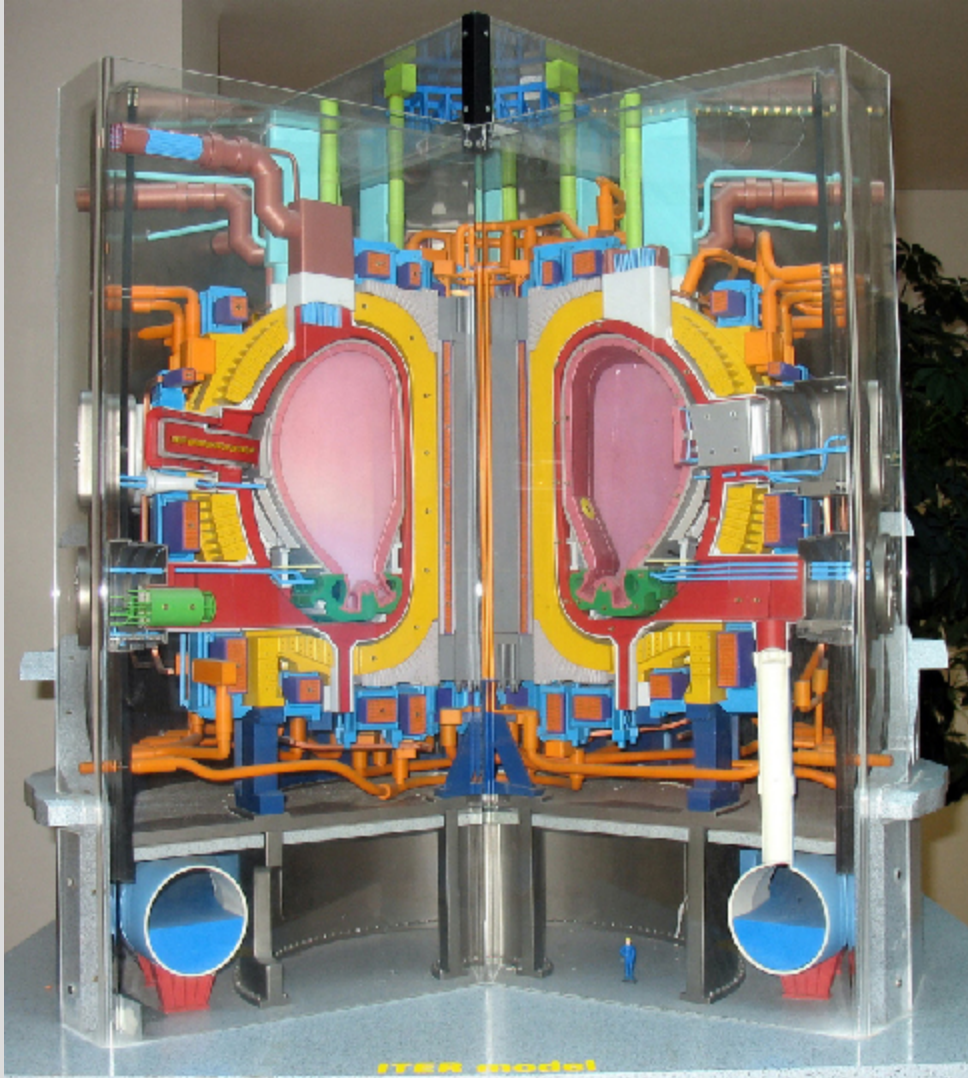
Note:**Fusion on Earth**

Wouldn't it be wonderful if we could duplicate the Sun's energy mechanism in a controlled way on Earth? (We have already duplicated it in an uncontrolled way in hydrogen bombs, but we hope our storehouses of these will never be used.) Fusion energy would have many advantages: it would use hydrogen (or deuterium, which is heavy hydrogen) as fuel, and there is abundant hydrogen in Earth's lakes and oceans. Water is much more evenly distributed around the world than oil or uranium, meaning that a few countries would no longer hold an energy advantage over the others. And unlike fission, which leaves dangerous byproducts, the nuclei that result from fusion are perfectly safe.

The problem is that, as we saw, it takes extremely high temperatures for nuclei to overcome their electrical repulsion and undergo fusion. When the first hydrogen bombs were exploded in tests in the 1950s, the "fuses" to get them hot enough were fission bombs. Interactions at such temperatures are difficult to sustain and control. To make fusion power on Earth, after all, we have to do what the Sun does: produce temperatures and pressures high enough to get hydrogen nuclei on intimate terms with one another. The European Union, the United States, South Korea, Japan, China, Russia, Switzerland, and India are collaborating on the International Thermonuclear Experimental Reactor (ITER), a project to demonstrate the feasibility of controlled fusion ([\[link\]](#)). The facility is being built in France. Construction will require over 10,000,000 components and 2000 workers for assembly. The date for the start of operations is yet to be determined. ITER is based on the Tokamak design, in which a large doughnut-shaped container is surrounded by superconducting magnets to confine and control the hydrogen nuclei in a strong magnetic field. Previous fusion experiments have produced about 15 million watts of energy, but only for a second or two, and they have required 100 million watts to produce the

conditions necessary to achieve fusion. The goal of ITER is to build the first fusion device capable of producing 500 million watts of fusion energy for up to 1000 seconds. The challenge is keeping the deuterium and tritium—which will participate in fusion reactions—hot enough and dense enough, for a long enough time to produce energy.

ITER Design.



The bright yellow areas in this model show where the superconducting magnets will circle the chamber within which fusion will take place. A huge magnet will keep the charged nuclei of heavy hydrogen confined. The goal is to produce 500 megawatts of

energy. (credit: modification of work by Stephan Mosel)

Key Concepts and Summary

Solar energy is produced by interactions of particles—that is, protons, neutrons, electrons, positrons, and neutrinos. Specifically, the source of the Sun’s energy is the fusion of hydrogen to form helium. The series of reactions required to convert hydrogen to helium is called the proton-proton chain. A helium atom is about 0.71% less massive than the four hydrogen atoms that combine to form it, and that lost mass is converted to energy (with the amount of energy given by the formula $E = mc^2$).

Glossary

fission

breaking up of heavier atomic nuclei into lighter ones

fusion

building up of heavier atomic nuclei from lighter ones

neutrino

fundamental particle that has no charge and a mass that is tiny relative to an electron; it rarely interacts with ordinary matter and comes in three different types

positron

particle with the same mass as an electron, but positively charged

proton-proton chain

series of thermonuclear reactions by which nuclei of hydrogen are built up into nuclei of helium

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Describe the state of equilibrium of the Sun
- Understand the energy balance of the Sun
- Explain how energy moves outward through the Sun
- Describe the structure of the solar interior

Fusion of protons can occur in the center of the Sun only if the temperature exceeds 12 million K. How do we know that the Sun is actually this hot? To determine what the interior of the Sun might be like, it is necessary to resort to complex calculations. Since we can't see the interior of the Sun, we have to use our understanding of physics, combined with what we see at the surface, to construct a mathematical model of what must be happening in the interior. Astronomers use observations to build a computer program containing everything they think they know about the physical processes going on in the Sun's interior. The computer then calculates the temperature and pressure at every point inside the Sun and determines what nuclear reactions, if any, are taking place. For some calculations, we can use observations to determine whether the computer program is producing results that match what we see. In this way, the program evolves with ever-improving observations.

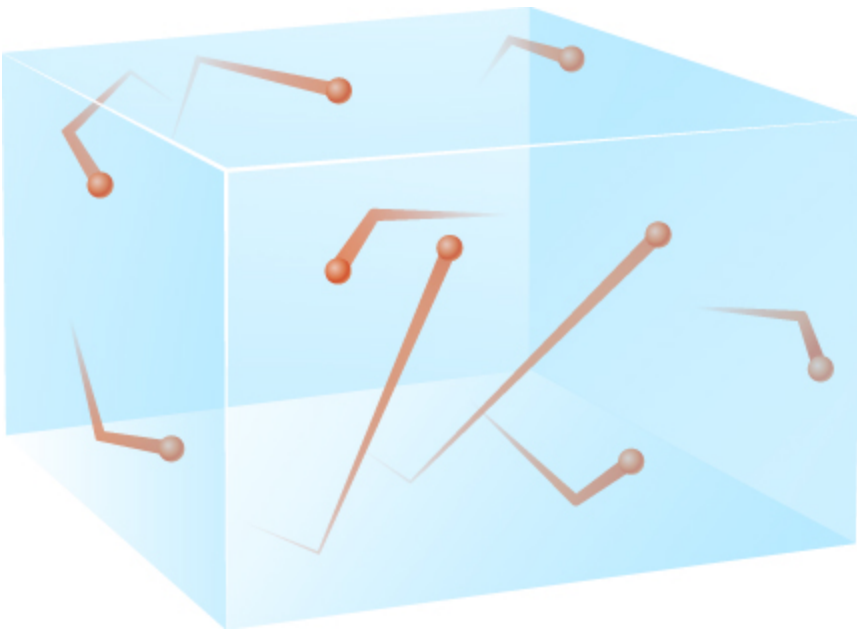
The computer program can also calculate how the Sun will change with time. After all, the Sun must change. In its center, the Sun is slowly depleting its supply of hydrogen and creating helium instead. Will the Sun

get hotter? Cooler? Larger? Smaller? Brighter? Fainter? Ultimately, the changes in the center could be catastrophic, since eventually all the hydrogen fuel hot enough for fusion will be exhausted. Either a new source of energy must be found, or the Sun will cease to shine. We will describe the ultimate fate of the Sun in later chapters. For now, let's look at some of the things we must teach the computer about the Sun in order to carry out such calculations.

The Sun Is a Plasma

The Sun is so hot that all of the material in it is in the form of an ionized gas, called a plasma. Plasma acts much like a hot gas, which is easier to describe mathematically than either liquids or solids. The particles that constitute a gas are in rapid motion, frequently colliding with one another. This constant bombardment is the *pressure* of the gas ([\[link\]](#)).

Gas Pressure.



The particles in a gas are in rapid motion and produce pressure through collisions with the surrounding material. Here, particles are shown bombarding the sides of an imaginary container.

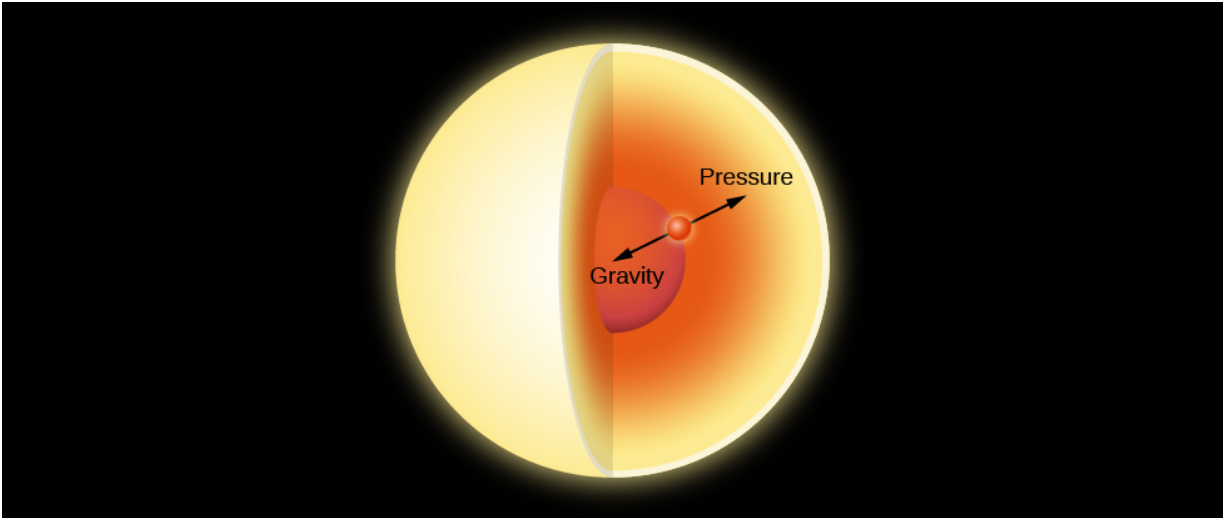
More particles within a given volume of gas produce more pressure because the combined impact of the moving particles increases with their number. The pressure is also greater when the molecules or atoms are moving faster. Since the molecules move faster when the temperature is hotter, higher temperatures produce higher pressure.

The Sun Is Stable

The Sun, like the majority of other stars, is stable; it is neither expanding nor contracting. Such a star is said to be in a condition of *equilibrium*. All the forces within it are balanced, so that at each point within the star, the temperature, pressure, density, and so on are maintained at constant values. We will see in later chapters that even these stable stars, including the Sun, are changing as they evolve, but such evolutionary changes are so gradual that, for all intents and purposes, the stars are still in a state of equilibrium at any given time.

The mutual gravitational attraction between the masses of various regions within the Sun produces tremendous forces that tend to collapse the Sun toward its center. Yet we know from the history of Earth that the Sun has been emitting roughly the same amount of energy for billions of years, so clearly it has managed to resist collapse for a very long time. The gravitational forces must therefore be counterbalanced by some other force. That force is due to the pressure of gases within the Sun ([link](#)). Calculations show that, in order to exert enough pressure to prevent the Sun from collapsing due to the force of gravity, the gases at its center must be maintained at a temperature of 15 million K. Think about what this tells us. Just from the fact that the Sun is not contracting, we can conclude that its temperature must indeed be high enough at the center for protons to undergo fusion.

Hydrostatic Equilibrium.



In the interior of a star, the inward force of gravity is exactly balanced at each point by the outward force of gas pressure.

The Sun maintains its stability in the following way. If the internal pressure in such a star were not great enough to balance the weight of its outer parts, the star would collapse somewhat, contracting and building up the pressure inside. On the other hand, if the pressure were greater than the weight of the overlying layers, the star would expand, thus decreasing the internal pressure. Expansion would stop, and equilibrium would again be reached when the pressure at every internal point equaled the weight of the stellar layers above that point. An analogy is an inflated balloon, which will expand or contract until an equilibrium is reached between the pressure of the air inside and outside. The technical term for this condition is **hydrostatic equilibrium**. Stable stars are all in hydrostatic equilibrium; so are the oceans of Earth as well as Earth's atmosphere. The air's own pressure keeps it from falling to the ground.

The Sun Is Not Cooling Down

As everyone who has ever left a window open on a cold winter night knows, heat always flows from hotter to cooler regions. As energy filters outward toward the surface of a star, it must be flowing from inner, hotter

regions. The temperature cannot ordinarily get cooler as we go inward in a star, or energy would flow in and heat up those regions until they were at least as hot as the outer ones. Scientists conclude that the temperature is highest at the center of a star, dropping to lower and lower values toward the stellar surface. (The high temperature of the Sun's chromosphere and corona may therefore appear to be a paradox. But remember from [The Sun: A Garden-Variety Star](#) that these high temperatures are maintained by magnetic effects, which occur in the Sun's atmosphere.)

The outward flow of energy through a star robs it of its internal heat, and the star would cool down if that energy were not replaced. Similarly, a hot iron begins to cool as soon as it is unplugged from its source of electric energy. Therefore, a source of fresh energy must exist within each star. In the Sun's case, we have seen that this energy source is the ongoing fusion of hydrogen to form helium.

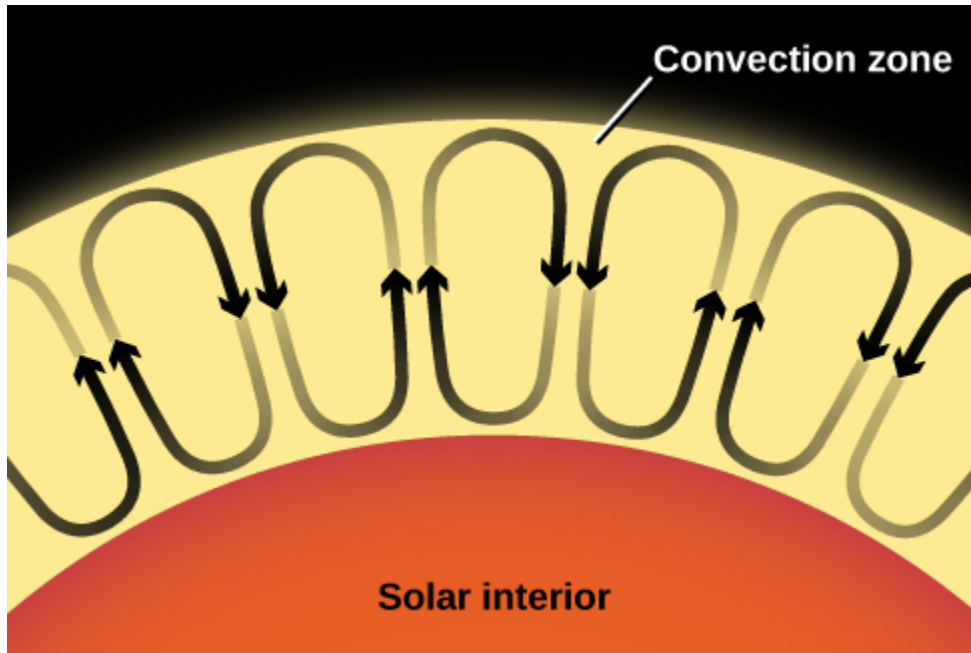
Heat Transfer in a Star

Since the nuclear reactions that generate the Sun's energy occur deep within it, the energy must be transported from the center of the Sun to its surface—where we see it in the form of both heat and light. There are three ways in which energy can be transferred from one place to another. In **conduction**, atoms or molecules pass on their energy by colliding with others nearby. This happens, for example, when the handle of a metal spoon heats up as you stir a cup of hot coffee. In **convection**, currents of warm material rise, carrying their energy with them to cooler layers. A good example is hot air rising from a fireplace. In **radiation**, energetic photons move away from hot material and are absorbed by some material to which they convey some or all of their energy. You can feel this when you put your hand close to the coils of an electric heater, allowing infrared photons to heat up your hand. Conduction and convection are both important in the interiors of planets. In stars, which are much more transparent, radiation and convection are important, whereas conduction can usually be ignored.

Stellar *convection* occurs as currents of hot gas flow up and down through the star ([\[link\]](#)). Such currents travel at moderate speeds and do not upset the overall stability of the star. They don't even result in a net transfer of

mass either inward or outward because, as hot material rises, cool material falls and replaces it. This results in a convective circulation of rising and falling cells as seen in [\[link\]](#). In much the same way, heat from a fireplace can stir up air currents in a room, some rising and some falling, without driving any air into or out the room. Convection currents carry heat very efficiently outward through a star. In the Sun, convection turns out to be important in the central regions and near the surface.

Convection.



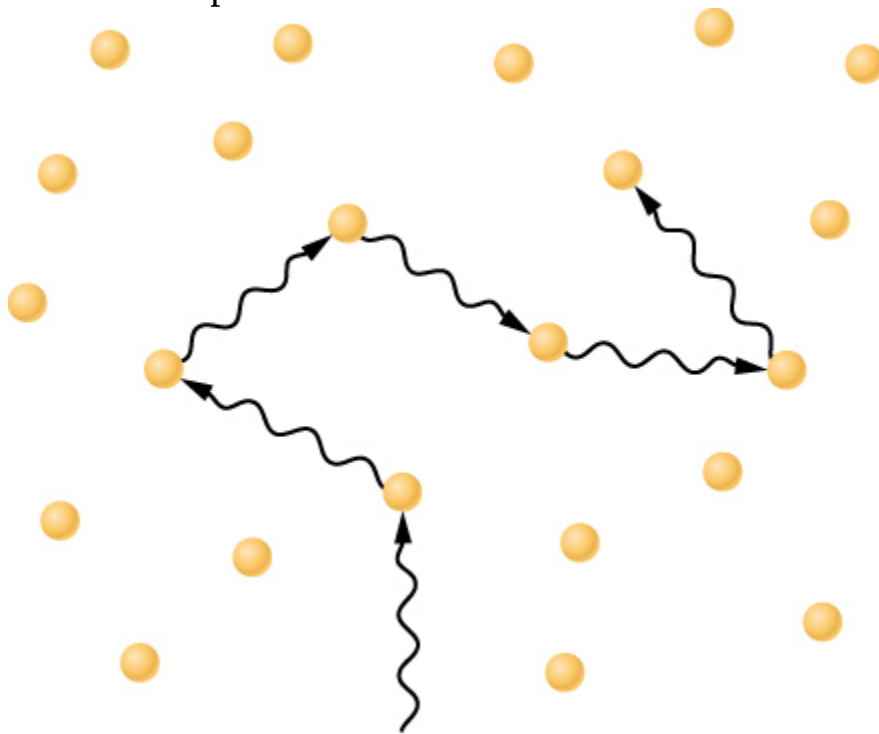
Rising convection currents carry heat from the Sun's interior to its surface, whereas cooler material sinks downward. Of course, nothing in a real star is as simple as diagrams in textbooks suggest.

Unless convection occurs, the only significant mode of energy transport through a star is by electromagnetic radiation. Radiation is not an efficient means of energy transport in stars because gases in stellar interiors are very opaque, that is, a photon does not go far (in the Sun, typically about 0.01 meter) before it is absorbed. (The processes by which atoms and ions can interrupt the outward flow of photons—such as becoming ionized—were discussed in the section on the [Formation of Spectral Lines](#).) The absorbed

energy is always reemitted, but it can be reemitted in any direction. A photon absorbed when traveling outward in a star has almost as good a chance of being radiated back toward the center of the star as toward its surface.

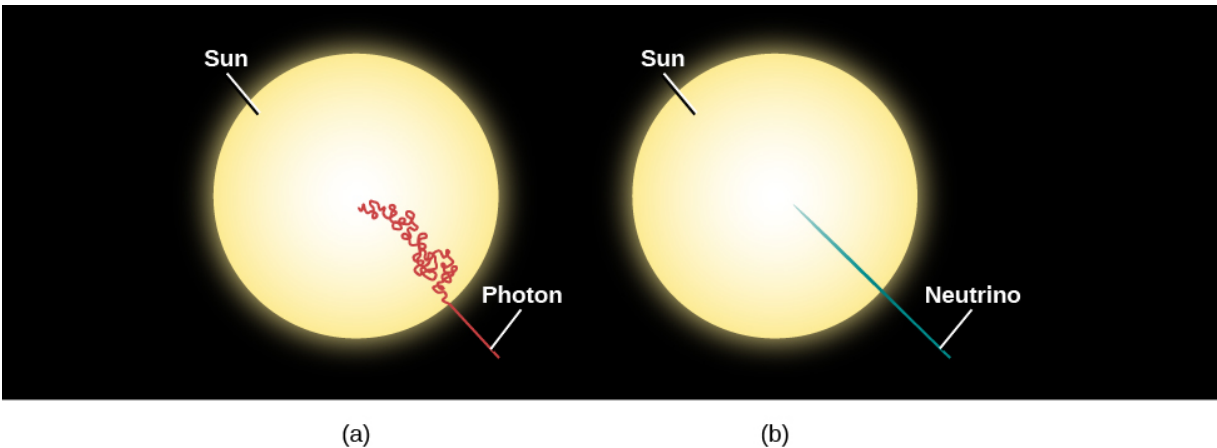
A particular quantity of energy, therefore, zigzags around in an almost random manner and takes a long time to work its way from the center of a star to its surface ([\[link\]](#)). Estimates are somewhat uncertain, but in the Sun, as we saw, the time required is probably between 100,000 and 1,000,000 years. If the photons were not absorbed and reemitted along the way, they would travel at the speed of light and could reach the surface in a little over 2 seconds, just as neutrinos do ([\[link\]](#)).

Photons Deep in the Sun.



A photon moving through the dense gases in the solar interior travels only a short distance before it interacts with one of the surrounding atoms. The resulting photon usually has a lower energy after each interaction and may then travel in any random direction.

Photon and Neutrino Paths in the Sun.



(a) Because photons generated by fusion reactions in the solar interior travel only a short distance before being absorbed or scattered by atoms and sent off in random directions, estimates are that it takes between 100,000 and 1,000,000 years for energy to make its way from the center of the Sun to its surface. (b) In contrast, neutrinos do not interact with matter but traverse straight through the Sun at the speed of light, reaching the surface in only a little more than 2 seconds.

Note:

Heat Transfer and Cooking

The three ways that heat energy moves from higher-temperature regions to cooler regions are all used in cooking, and this is important to all of us who enjoy making or eating food.

Conduction is heat transfer by physical contact during which the energetic motion of particles in one region spread to other regions and even to adjacent objects in close contact. A tasty example of this is cooking a steak on a hot iron skillet. When a flame makes the bottom of a skillet hot, the particles in it vibrate actively and collide with neighboring particles, spreading the heat energy throughout the skillet (the ability to spread heat uniformly is a key criterion for selecting materials for cookware). A steak sitting on the surface of the skillet picks up heat energy by the particles in

the surface of the skillet colliding with particles on the surface of the steak. Many cooks will put a little oil on the pan, and this layer of oil, besides preventing sticking, increases heat transfer by filling in gaps and increasing the contact surface area.

Convection is heat transfer by the motion of matter that rises because it is hot and less dense. Heating a fluid makes it expand, which makes it less dense, so it rises. An oven is a great example of this: the fire is at the bottom of the oven and heats the air down there, causing it to expand (becoming less dense), so it rises up to where the food is. The rising hot air carries the heat from the fire to the food by convection. This is how conventional ovens work. You may also be familiar with convection ovens that use a fan to circulate hot air for more even cooking. A scientist would object to that name because normal non-fan ovens that rely on hot air rising to circulate the heat are convection ovens; technically, the ovens that use fans to help move heat are “advection” ovens. (You may not have heard about this because the scientists who complain loudly about misusing the terms convection and advection don’t get out much.)

Radiation is the transfer of heat energy by electromagnetic radiation.

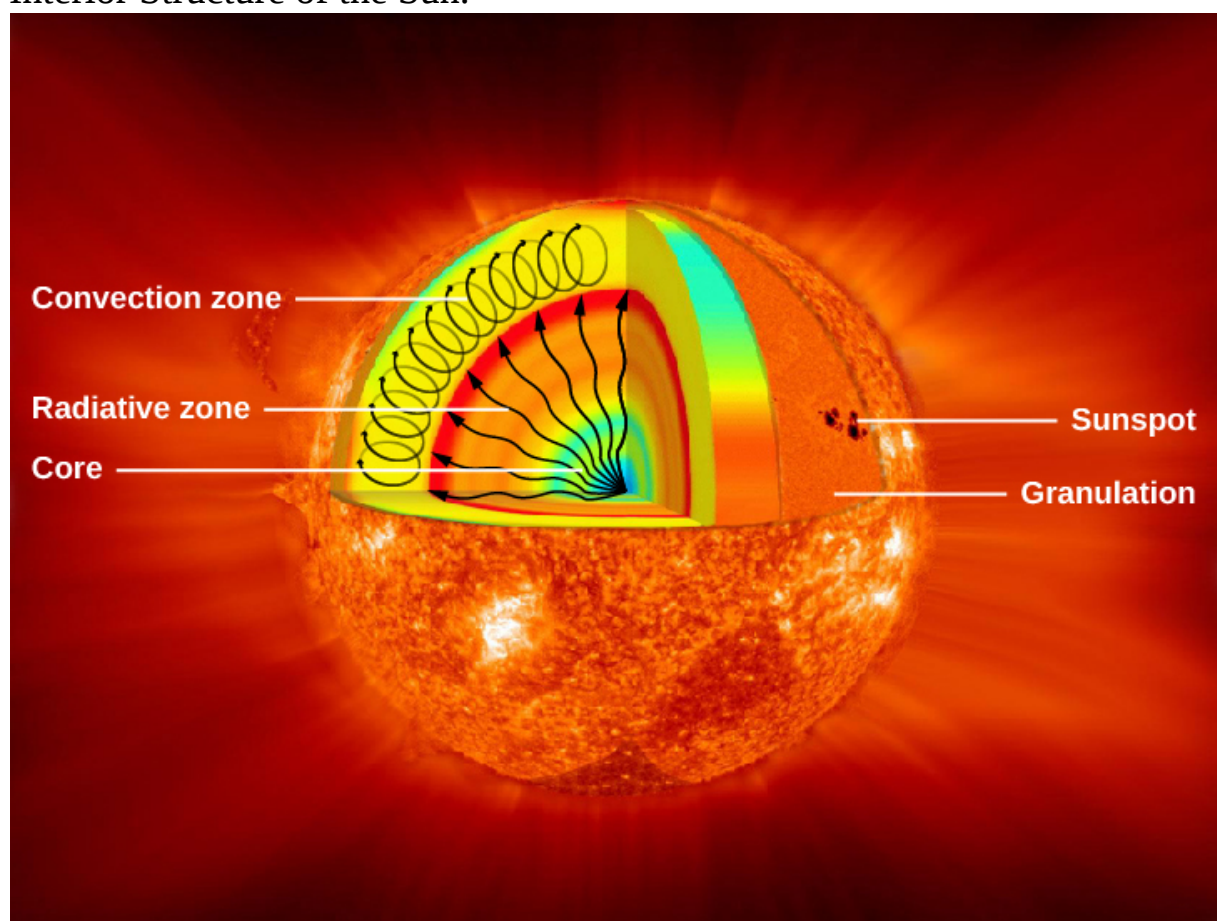
Although microwave ovens are an obvious example of using radiation to heat food, a simpler example is a toy oven. Toy ovens are powered by a very bright light bulb. The child-chefs prepare a mix for brownies or cookies, put it into a tray, and place it in the toy oven under the bright light bulb. The light and heat from the bulb hit the brownie mix and cook it. If you have ever put your hand near a bright light, you have undoubtedly noticed your hand getting warmed by the light.

Model Stars

Scientists use the principles we have just described to calculate what the Sun’s interior is like. These physical ideas are expressed as mathematical equations that are solved to determine the values of temperature, pressure, density, the efficiency with which photons are absorbed, and other physical quantities throughout the Sun. The solutions obtained, based on a specific set of physical assumptions, provide a theoretical model for the interior of the Sun.

[\[link\]](#) schematically illustrates the predictions of a theoretical model for the Sun's interior. Energy is generated through fusion in the core of the Sun, which extends only about one-quarter of the way to the surface but contains about one-third of the total mass of the Sun. At the center, the temperature reaches a maximum of approximately 15 million K, and the density is nearly 150 times that of water. The energy generated in the core is transported toward the surface by radiation until it reaches a point about 70% of the distance from the center to the surface. At this point, convection begins, and energy is transported the rest of the way, primarily by rising columns of hot gas.

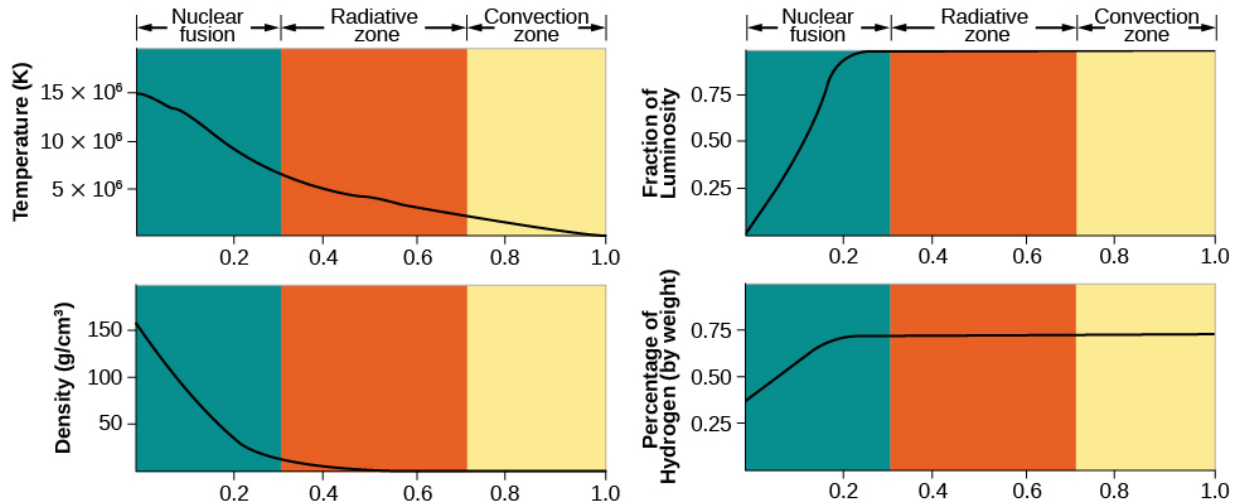
Interior Structure of the Sun.



Energy is generated in the core by the fusion of hydrogen to form helium. This energy is transmitted outward by radiation—that is, by the absorption and reemission of photons. In the outermost layers, energy is transported mainly by convection. (credit: modification of work by NASA/Goddard)

[\[link\]](#) shows how the temperature, density, rate of energy generation, and composition vary from the center of the Sun to its surface.

Interior of the Sun.



Diagrams showing how temperature, density, rate of energy generation, and the percentage (by mass) abundance of hydrogen vary inside the Sun. The horizontal scale shows the fraction of the Sun's radius: the left edge is the very center, and the right edge is the visible surface of the Sun, which is called the photosphere.

Key Concepts and Summary

Even though we cannot see inside the Sun, it is possible to calculate what its interior must be like. As input for these calculations, we use what we know about the Sun. It is made entirely of hot gas. Apart from some very tiny changes, the Sun is neither expanding nor contracting (it is in hydrostatic equilibrium) and puts out energy at a constant rate. Fusion of hydrogen occurs in the center of the Sun, and the energy generated is carried to the surface by radiation and then convection. A solar model describes the structure of the Sun's interior. Specifically, it describes how

pressure, temperature, mass, and luminosity depend on the distance from the center of the Sun.

Glossary

conduction

process by which heat is directly transmitted through a substance when there is a difference of temperature between adjoining regions caused by atomic or molecular collisions

convection

movement caused within a gas or liquid by the tendency of hotter, and therefore less dense material, to rise and colder, denser material to sink under the influence of gravity, which consequently results in transfer of heat

hydrostatic equilibrium

balance between the weights of various layers, as in a star or Earth's atmosphere, and the pressures that support them

radiation

emission of energy as electromagnetic waves or photons also the transmitted energy itself

The Solar Interior: Observations

LEARNING OBJECTIVES

By the end of this section, you will be able to:

- Explain how the Sun pulsates
- Explain what helioseismology is and what it can tell us about the solar interior
- Discuss how studying neutrinos from the Sun has helped understand neutrinos

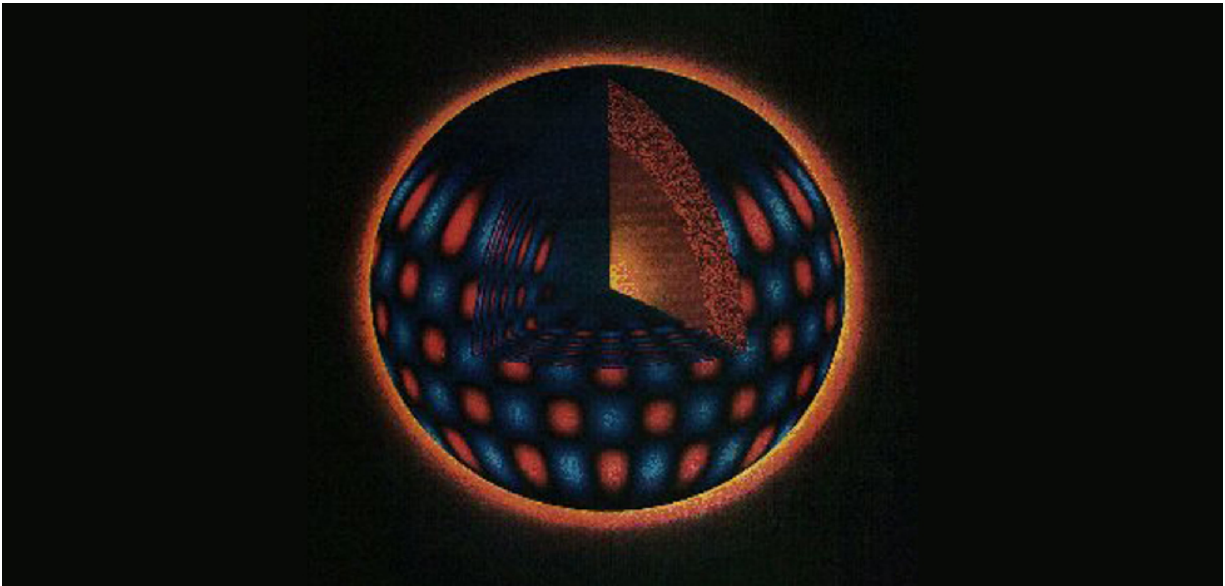
Recall that when we observe the Sun's photosphere (the surface layer we see from the outside), we are not seeing very deeply into our star, certainly not into the regions where energy is generated. That's why the title of this section—observations of the solar interior—should seem very surprising. However, astronomers have indeed devised two types of measurements that can be used to obtain information about the inner parts of the Sun. One technique involves the analysis of tiny changes in the motion of small regions at the Sun's surface. The other relies on the measurement of the neutrinos emitted by the Sun.

Solar Pulsations

Astronomers discovered that the Sun pulsates—that is, it alternately expands and contracts—just as your chest expands and contracts as you breathe. This pulsation is very slight, but it can be detected by measuring the *radial velocity* of the solar surface—the speed with which it moves toward or away from us. The velocities of small regions on the Sun are

observed to change in a regular way, first toward Earth, then away, then toward, and so on. It is as if the Sun were “breathing” through thousands of individual lungs, each having a size in the range of 4000 to 15,000 kilometers, each fluctuating back and forth ([link](#)).

Oscillations in the Sun.



New observational techniques permit astronomers to measure small differences in velocity at the Sun's surface to infer what the deep solar interior is like. In this computer simulation, red shows surface regions that are moving away from the observer (inward motion); blue marks regions moving toward the observer (outward motion). Note that the velocity changes penetrate deep into the Sun's interior. (credit: modification of work by GONG, NOAO)

The typical velocity of one of the oscillating regions on the Sun is only a few hundred meters per second, and it takes about 5 minutes to complete a full cycle from maximum to minimum velocity and back again. The change in the size of the Sun measured at any given point is no more than a few kilometers.

The remarkable thing is that these small velocity variations can be used to determine what the interior of the Sun is like. The motion of the Sun's

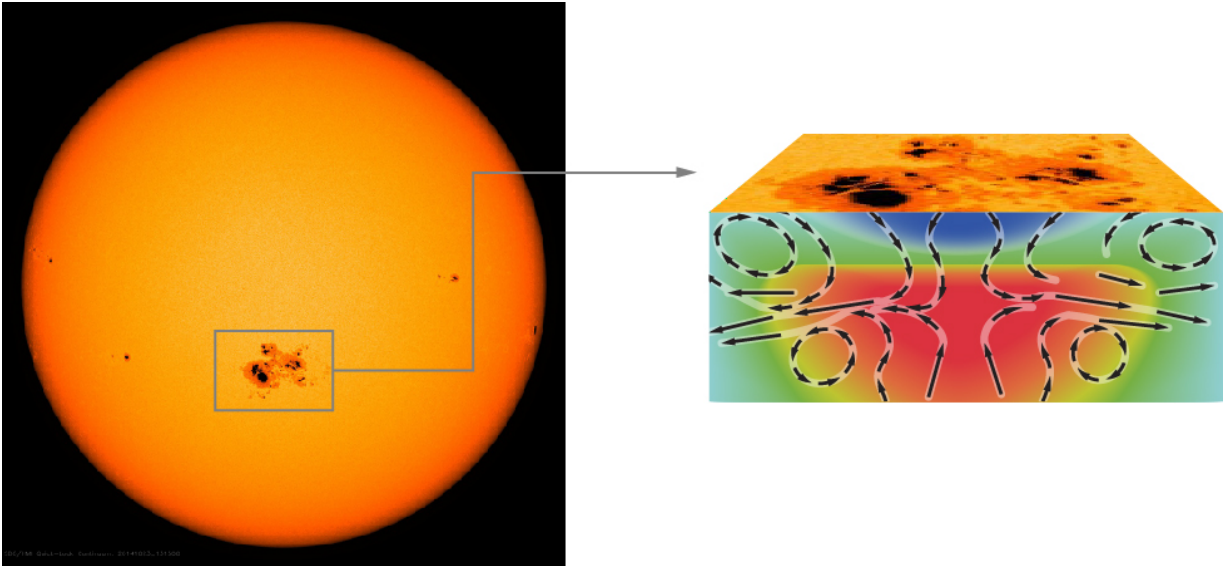
surface is caused by waves that reach it from deep in the interior. Study of the amplitude and cycle length of velocity changes provides information about the temperature, density, and composition of the layers through which the waves passed before they reached the surface. The situation is somewhat analogous to the use of seismic waves generated by earthquakes to infer the properties of Earth's interior. For this reason, studies of solar oscillations (back-and-forth motions) are referred to as **helioseismology**.

It takes a little over an hour for waves to traverse the Sun from center to surface, so the waves, like neutrinos, provide information about what the solar interior is like at the present time. In contrast, remember that the sunlight we see today emerging from the Sun was actually generated in the core several hundred thousand years ago.

Helioseismology has shown that convection extends inward from the surface 30% of the way toward the center; we have used this information in drawing [\[link\]](#). Pulsation measurements also show that the *differential rotation* that we see at the Sun's surface, with the fastest rotation occurring at the equator, persists down through the convection zone. Below the convection zone, however, the Sun, even though it is gaseous throughout, rotates as if it were a solid body like a bowling ball. Another finding from helioseismology is that the abundance of helium inside the Sun, except in the center where nuclear reactions have converted hydrogen into helium, is about the same as at its surface. That result is important to astronomers because it means we are correct when we use the abundance of the elements measured in the solar atmosphere to construct models of the solar interior.

Helioseismology also allows scientists to look beneath a sunspot and see how it works. In [The Sun: A Garden-Variety Star](#), we said that sunspots are cool because strong magnetic fields block the outward flow of energy. [\[link\]](#) shows how gas moves around underneath a sunspot. Cool material from the sunspot flows downward, and material surrounding the sunspot is pulled inward, carrying magnetic field with it and thus maintaining the strong field that is necessary to form a sunspot. As the new material enters the sunspot region, it too cools, becomes denser, and sinks, thus setting up a self-perpetuating cycle that can last for weeks.

Sunspot Structure.



This drawing shows our new understanding, from helioseismology, of what lies beneath a sunspot. The black arrows show the direction of the flow of material. The intense magnetic field associated with the sunspot stops the upward flow of hot material and creates a kind of plug that blocks the hot gas. As the material above the plug cools (shown in blue), it becomes denser and plunges inward, drawing more gas and more magnetic field behind it into the spot. The concentrated magnetic field causes more cooling, thereby setting up a self-perpetuating cycle that allows a spot to survive for several weeks. Since the plug keeps hot material from flowing up into the sunspot, the region below the plug, represented by red in this picture, becomes hotter. This material flows sideways and then upward, eventually reaching the solar surface in the area surrounding the sunspot. (credit: modification of work by NASA, SDO)

The downward-flowing cool material acts as a kind of plug that block the upward flow of hot material, which is then diverted sideways and eventually reaches the solar surface in the region around the sunspot. This outward flow of hot material accounts for the paradox that we described in [The Sun: A Garden-Variety Star](#)—namely, that the Sun emits slightly more energy when more of its surface is covered by cool sunspots.

Helioseismology has become an important tool for predicting solar storms that might impact Earth. Active regions can appear and grow large in only a few days. The solar rotation period is about 28 days. Therefore, regions capable of producing solar flares and coronal mass ejections can develop on the far side of the Sun, where, for a long time, we couldn't see them directly.

Fortunately, we now have space telescopes monitoring the Sun from all angles, so we know if there are sunspots forming on the opposite side of the Sun. Moreover, sound waves travel slightly faster in regions of high magnetic field, and waves generated in active regions traverse the Sun about 6 seconds faster than waves generated in quiet regions. By detecting this subtle difference, scientists can provide warnings of a week or more to operators of electric utilities and satellites about when a potentially dangerous active region might rotate into view. With this warning, it is possible to plan for disruptions, put key instruments into safe mode, or reschedule spacewalks in order to protect astronauts.

Solar Neutrinos

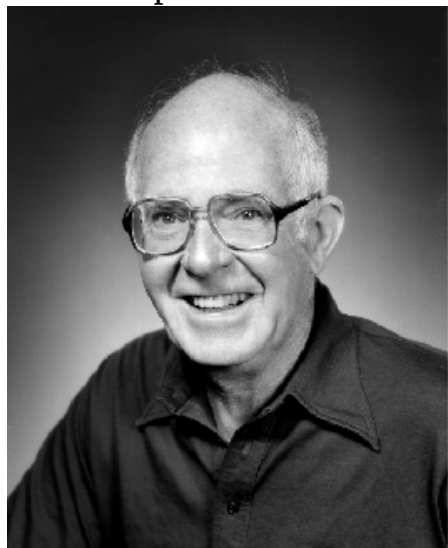
The second technique for obtaining information about the Sun's interior involves the detection of a few of those elusive neutrinos created during nuclear fusion. Recall from our earlier discussion that neutrinos created in the center of the Sun make their way directly out of the Sun and travel to Earth at nearly the speed of light. As far as neutrinos are concerned, the Sun is transparent.

About 3% of the total energy generated by nuclear fusion in the Sun is carried away by neutrinos. So many protons react and form neutrinos inside the Sun's core that, scientists calculate, 35 million billion (3.5×10^{16}) solar neutrinos pass through each square meter of Earth's surface every second. If we can devise a way to detect even a few of these solar neutrinos, then we can obtain information directly about what is going on in the center of the Sun. Unfortunately for those trying to "catch" some neutrinos, Earth and everything on it are also nearly transparent to passing neutrinos, just like the Sun.

On very, very rare occasions, however, one of the billions and billions of solar neutrinos will interact with another atom. The first successful detection of solar neutrinos made use of cleaning fluid (C_2Cl_4), which is the least expensive way to get a lot of chlorine atoms together. The nucleus of a chlorine (Cl) atom in the cleaning fluid can be turned into a radioactive argon nucleus by an interaction with a neutrino. Because the argon is radioactive, its presence can be detected. However, since the interaction of a neutrino with chlorine happens so rarely, a huge amount of chlorine is needed.

Raymond Davis, Jr. ([\[link\]](#)) and his colleagues at Brookhaven National Laboratory, placed a tank containing nearly 400,000 liters of cleaning fluid 1.5 kilometers beneath Earth's surface in a gold mine at Lead, South Dakota. A mine was chosen so that the surrounding material of Earth would keep cosmic rays (high-energy particles from space) from reaching the cleaning fluid and creating false signals. (Cosmic-ray particles are stopped by thick layers of Earth, but neutrinos find them of no significance.) Calculations show that solar neutrinos should produce about one atom of radioactive argon in the tank each day.

Davis Experiment.



(a)



(b)

(a) Raymond Davis received the Nobel Prize in physics in 2002. (b) Davis' experiment at the bottom of an abandoned gold mine first revealed problems with our understanding of neutrinos. (credit a:

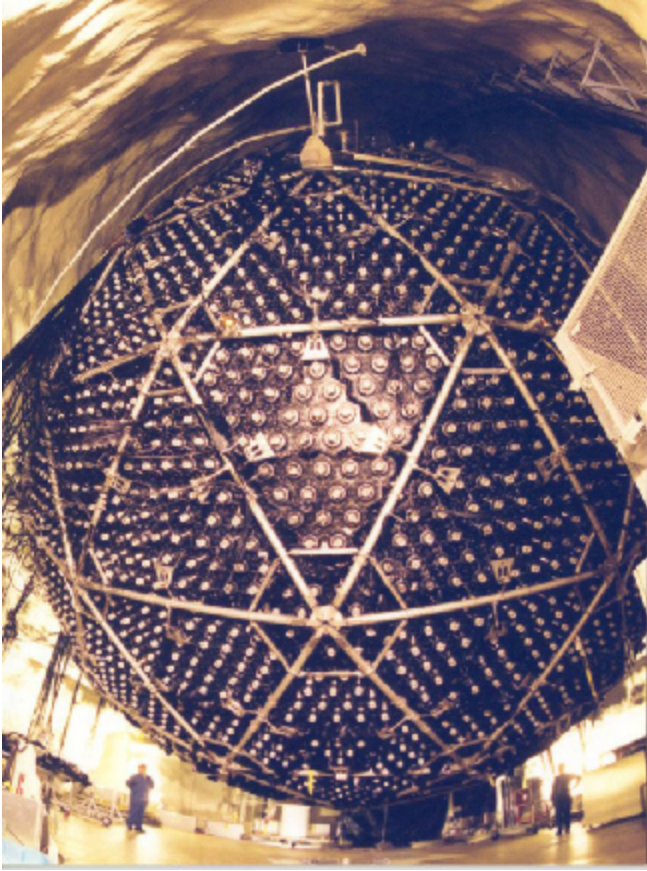
modification of work by Brookhaven National Laboratory; credit b:
modification of work by the United States Department of Energy)

This was an amazing project: they counted argon atoms about once per month—and remember, they were looking for a tiny handful of argon atoms in a massive tank of chlorine atoms. When all was said and done, Davis’ experiment, begun in 1970, detected only about one-third as many neutrinos as predicted by solar models! This was a shocking result because astronomers thought they had a pretty good understanding of both neutrinos and the Sun’s interior. For many years, astronomers and physicists wrestled with Davis’ results, trying to find a way out of the dilemma of the “missing” neutrinos.

Eventually Davis’ result was explained by the surprising discovery that there are actually three types of neutrinos. Solar fusion produces only one type of neutrino, the so-called electron neutrino, and the initial experiments to detect solar neutrinos were designed to detect this one type. Subsequent experiments showed that these neutrinos change to a different type during their journey from the center of the Sun through space to Earth in a process called *neutrino oscillation*.

An experiment, conducted at the Sudbury Neutrino Observatory in Canada, was the first one designed to capture all three types of neutrinos ([\[link\]](#)). The experiment was located in a mine 2 kilometers underground. The neutrino detector consisted of a 12-meter-diameter transparent acrylic plastic sphere, which contained 1000 metric tons of heavy water. Remember that an ordinary water nucleus contains two hydrogen atoms and one oxygen atom. Heavy water instead contains two deuterium atoms and one oxygen atom, and incoming neutrinos can occasionally break up the loosely bound proton and neutron that make up the deuterium nucleus. The sphere of heavy water was surrounded by a shield of 1700 metric tons of very pure water, which in turn was surrounded by 9600 photomultipliers, devices that detect flashes of light produced after neutrinos interact with the heavy water.

Sudbury Neutrino Detector.



The 12-meter sphere of the Sudbury Neutrino Detector lies more than 2 kilometers underground and holds 1000 metric tons of heavy water. (credit: A.B. McDonald (Queen's University) et al., The Sudbury Neutrino Observatory Institute)

To the enormous relief of astronomers who make models of the Sun, the Sudbury experiment detected about 1 neutrino per hour and has shown that the *total* number of neutrinos reaching the heavy water is just what solar models predict. Only one-third of these, however, are electron neutrinos. It appears that two-thirds of the electron neutrinos produced by the Sun transform themselves into one of the other types of neutrinos as they make

their way from the core of the Sun to Earth. This is why the earlier experiments saw only one-third the number of neutrinos expected.

Although it is not intuitively obvious, such neutrino oscillations can happen only if the mass of the electron neutrino is not zero. Other experiments indicate that its mass is tiny (even compared to the electron). The 2015 Nobel Prize in physics was awarded to researchers Takaaki Kajita and Arthur B. McDonald for their work establishing the changeable nature of neutrinos. (Raymond Davis shared the 2002 Nobel Prize with Japan's Masatoshi Koshiba for the experiments that led to our understanding of the neutrino problem in the first place.) But the fact that the neutrino has mass at all has deep implications for both physics and astronomy. For example, we will look at the role that neutrinos play in the inventory of the mass of the universe in [The Big Bang](#).

The Borexino experiment, an international experiment conducted in Italy, detected neutrinos coming from the Sun that were identified as coming from different reactions. Whereas the p-p chain is the reaction producing most of the Sun's energy, it is not the only nuclear reaction occurring in the Sun's core. There are side reactions involving nuclei of such elements as beryllium and boron. By probing the number of neutrinos that come from each reaction, the Borexino experiment has helped us confirm in detail our understanding of nuclear fusion in the Sun. In 2014, the Borexino experiment also identified neutrinos that were produced by the first step in the p-p chain, confirming the models of solar astronomers.

It's amazing that a series of experiments that began with enough cleaning fluid to fill a swimming pool brought down the shafts of an old gold mine is now teaching us about the energy source of the Sun and the properties of matter! This is a good example of how experiments in astronomy and physics, coupled with the best theoretical models we can devise, continue to lead to fundamental changes in our understanding of nature.

Key Concepts and Summary

Studies of solar oscillations (helioseismology) and neutrinos can provide observational data about the Sun's interior. The technique of

helioseismology has so far shown that the composition of the interior is much like that of the surface (except in the core, where some of the original hydrogen has been converted into helium), and that the convection zone extends about 30% of the way from the Sun's surface to its center. Helioseismology can also detect active regions on the far side of the Sun and provide better predictions of solar storms that may affect Earth. Neutrinos from the Sun call tell us about what is happening in the solar interior. A recent experiment has shown that solar models do predict accurately the number of electron neutrinos produced by nuclear reactions in the core of the Sun. However, two-thirds of these neutrinos are converted into different types of neutrinos during their long journey from the Sun to Earth, a result that also indicates that neutrinos are not massless particles.

For Further Exploration

Articles

Harvey, J. et al. "GONG: To See Inside Our Sun." *Sky & Telescope* (November 1987): 470.

Hathaway, D. "Journey to the Heart of the Sun." *Astronomy* (January 1995): 38.

Kennedy, J. "GONG: Probing the Sun's Hidden Heart." *Sky & Telescope* (October 1996): 20. A discussion on hydroseismology.

LoPresto, J. "Looking Inside the Sun." *Astronomy* (March 1989): 20. A discussion on hydroseismology.

McDonald, A. et al. "Solving the Solar Neutrino Problem." *Scientific American* (April 2003): 40. A discussion on how underground experiments with neutrino detectors helped explain the seeming absence of neutrinos from the Sun.

Trefil, J. "How Stars Shine." *Astronomy* (January 1998): 56.

Websites

Albert Einstein Online: <http://www.westegg.com/einstein/>.

Ghost Particle: <http://www.pbs.org/wgbh/nova/neutrino/>.

GONG Project Site: <http://gong.nso.edu/>.

Helioseismology: <http://solar-center.stanford.edu/about/helioseismology.html>.

Princeton Plasma Physics Lab: <http://www.pppl.gov/>.

Solving the Mystery of the Solar Neutrinos:
http://www.nobelprize.org/nobel_prizes/themes/physics/bahcall/.

Super Kamiokande Neutrino Mass Page: <http://www.ps.uci.edu/~superk/>.

Videos

Deep Secrets of the Neutrino: Physics Underground:
<https://www.youtube.com/watch?v=Ar9ydagYkYg>. 2010 Public Lecture by Peter Rowson at the Stanford Linear Accelerator Center (1:22:00).

The Elusive Neutrino and the Nature of Physics:
<https://www.youtube.com/watch?v=CBfUHzkcaHQ>. Panel at the 2014 World Science Festival (1:30:00).

The Ghost Particle: http://www.dailymotion.com/video/x20rn7s_nova-the-ghost-particle-discovery-science-universe-documentary_tv. 2006 NOVA episode (52:49).

Collaborative Group Activities

- A. In this chapter, we learned that meteorites falling into the Sun could not be the source of the Sun's energy because the necessary increase in

the mass of the Sun would lengthen Earth's orbital period by 2 seconds per year. Have your group discuss what effects this would cause for our planet and for us as the centuries went on.

- B. Solar astronomers can learn more about the Sun's interior if they can observe the Sun's oscillations 24 hours each day. This means that they cannot have their observations interrupted by the day/night cycle. Such an experiment, called the GONG (Global Oscillation Network Group) project, was first set up in the 1990s. To save money, this experiment was designed to make use of the minimum possible number of telescopes. It turns out that if the sites are selected carefully, the Sun can be observed all but about 10% of the time with only six observing stations. What factors do you think have to be taken into consideration in selecting the observing sites? Can your group suggest six general geographic locations that would optimize the amount of time that the Sun can be observed? Check your answer by looking at the GONG website.
- C. What would it be like if we actually manage to get controlled fusion on Earth to be economically feasible? If the hydrogen in *water* becomes the fuel for releasing enormous amounts of energy (instead of fossil fuels), have your group discuss how this affects the world economy and international politics. (Think of the role that oil and natural gas deposits now play on the world scene and in international politics.)
- D. Your group is a delegation sent to the city council of a small mining town to explain why the government is putting a swimming-pool-sized vat of commercial cleaning fluid down one of the shafts of an old gold mine. How would you approach this meeting? Assuming that the members of the city council do not have much science background, how would you explain the importance of the project to them? Suggest some visual aids you could use.
- E. When Raymond Davis first suggested his experiment in the underground gold mine, which had significant costs associated with it, some people said it wasn't worth the expense since we already understood the conditions and reactions in the core of the Sun. Yet his experiment led to a major change in our understanding of neutrinos and the physics of subatomic particles. Can your group think of other "expensive" experiments in astronomy that led to fundamental improvements in our understanding of nature?

Review Questions

Exercise:

Problem: How do we know the age of the Sun?

Exercise:

Problem:

Explain how we know that the Sun's energy is not supplied either by chemical burning, as in fires here on Earth, or by gravitational contraction (shrinking).

Exercise:

Problem:

What is the ultimate source of energy that makes the Sun shine?

Exercise:

Problem:

What are the formulas for the three steps in the proton-proton chain?

Exercise:

Problem:

How is a neutrino different from a neutron? List all the ways you can think of.

Exercise:

Problem:

Describe in your own words what is meant by the statement that the Sun is in hydrostatic equilibrium.

Exercise:

Problem:

Two astronomy students travel to South Dakota. One stands on Earth's surface and enjoys some sunshine. At the same time, the other descends into a gold mine where neutrinos are detected, arriving in time to detect the creation of a new radioactive argon nucleus. Although the photon at the surface and the neutrinos in the mine arrive at the same time, they have had very different histories. Describe the differences.

Exercise:**Problem:**

What do measurements of the number of neutrinos emitted by the Sun tell us about conditions deep in the solar interior?

Exercise:**Problem:**

Do neutrinos have mass? Describe how the answer to this question has changed over time and why.

Exercise:**Problem:**

Neutrinos produced in the core of the Sun carry energy to its exterior. Is the mechanism for this energy transport conduction, convection, or radiation?

Exercise:**Problem:**

What conditions are required before proton-proton chain fusion can start in the Sun?

Exercise:

Problem:

Describe the two main ways that energy travels through the Sun.

Thought Questions**Exercise:****Problem:**

Someone suggests that astronomers build a special gamma-ray detector to detect gamma rays produced during the proton-proton chain in the core of the Sun, just like they built a neutrino detector. Explain why this would be a fruitless effort.

Exercise:**Problem:**

Earth contains radioactive elements whose decay produces neutrinos. How might we use neutrinos to determine how these elements are distributed in Earth's interior?

Exercise:**Problem:**

The Sun is much larger and more massive than Earth. Do you think the average density of the Sun is larger or smaller than that of Earth? Write down your answer before you look up the densities. Now find the values of the densities elsewhere in this text. Were you right? Explain clearly the meanings of density and mass.

Exercise:**Problem:**

A friend who has not had the benefit of an astronomy course suggests that the Sun must be full of burning coal to shine as brightly as it does. List as many arguments as you can against this hypothesis.

Exercise:**Problem:**

Which of the following transformations is (are) fusion and which is (are) fission: helium to carbon, carbon to iron, uranium to lead, boron to carbon, oxygen to neon? (See [Appendix K](#) for a list of the elements.)

Exercise:**Problem:**

Why is a higher temperature required to fuse hydrogen to helium by means of the CNO cycle than is required by the process that occurs in the Sun, which involves only isotopes of hydrogen and helium?

Exercise:**Problem:**

Earth's atmosphere is in hydrostatic equilibrium. What this means is that the pressure at any point in the atmosphere must be high enough to support the weight of air above it. How would you expect the pressure on Mt. Everest to differ from the pressure in your classroom? Explain why.

Exercise:**Problem:**

Explain what it means when we say that Earth's oceans are in hydrostatic equilibrium. Now suppose you are a scuba diver. Would you expect the pressure to increase or decrease as you dive below the surface to a depth of 200 feet? Why?

Exercise:**Problem:**

What mechanism transfers heat away from the surface of the Moon? If the Moon is losing energy in this way, why does it not simply become colder and colder?

Exercise:**Problem:**

Suppose you are standing a few feet away from a bonfire on a cold fall evening. Your face begins to feel hot. What is the mechanism that transfers heat from the fire to your face? (Hint: Is the air between you and the fire hotter or cooler than your face?)

Exercise:**Problem:**

Give some everyday examples of the transport of heat by convection and by radiation.

Exercise:**Problem:**

Suppose the proton-proton cycle in the Sun were to slow down suddenly and generate energy at only 95% of its current rate. Would an observer on Earth see an immediate decrease in the Sun's brightness? Would she immediately see a decrease in the number of neutrinos emitted by the Sun?

Exercise:**Problem:**

Do you think that nuclear fusion takes place in the atmospheres of stars? Why or why not?

Exercise:

Problem: Why is fission not an important energy source in the Sun?

Exercise:

Problem:

Why do you suppose so great a fraction of the Sun's energy comes from its central regions? Within what fraction of the Sun's radius does practically all of the Sun's luminosity originate (see [\[link\]](#))? Within what radius of the Sun has its original hydrogen been partially used up? Discuss what relationship the answers to these questions bear to one another.

Exercise:**Problem:**

Explain how mathematical computer models allow us to understand what is going on inside of the Sun.

Figuring for Yourself**Exercise:****Problem:**

Estimate the amount of mass that is converted to energy when a proton combines with a deuterium nucleus to form ${}^3\text{He}$.

Exercise:**Problem:**

How much energy is released when a proton combines with a deuterium nucleus to produce ${}^3\text{He}$?

Exercise:**Problem:**

The Sun converts 4×10^9 kg of mass to energy every second. How many years would it take the Sun to convert a mass equal to the mass of Earth to energy?

Exercise:

Problem:

Assume that the mass of the Sun is 75% hydrogen and that all of this mass could be converted to energy according to Einstein's equation $E = mc^2$. How much total energy could the Sun generate? If m is in kg and c is in m/s, then E will be expressed in J. (The mass of the Sun is given in [Appendix E](#).)

Exercise:**Problem:**

In fact, the conversion of mass to energy in the Sun is not 100% efficient. As we have seen in the text, the conversion of four hydrogen atoms to one helium atom results in the conversion of about 0.02862 times the mass of a proton to energy. How much energy in joules does one such reaction produce? (See [Appendix E](#) for the mass of the hydrogen atom, which, for all practical purposes, is the mass of a proton.)

Exercise:**Problem:**

Now suppose that all of the hydrogen atoms in the Sun were converted into helium. How much total energy would be produced? (To calculate the answer, you will have to estimate how many hydrogen atoms are in the Sun. This will give you good practice with scientific notation, since the numbers involved are very large! See [Appendix C](#) for a review of scientific notation.)

Exercise:

Problem:

Models of the Sun indicate that only about 10% of the total hydrogen in the Sun will participate in nuclear reactions, since it is only the hydrogen in the central regions that is at a high enough temperature. Use the total energy radiated per second by the Sun, 3.8×10^{26} watts, alongside the exercises and information given here to estimate the lifetime of the Sun. (Hint: Make sure you keep track of the units: if the luminosity is the energy radiated per second, your answer will also be in seconds. You should convert the answer to something more meaningful, such as years.)

Exercise:**Problem:**

Show that the statement in the text is correct: namely, that roughly 600 million tons of hydrogen must be converted to helium in the Sun each second to explain its energy output. (Hint: Recall Einstein's most famous formula, and remember that for each kg of hydrogen, 0.0071 kg of mass is converted into energy.) How long will it be before 10% of the hydrogen is converted into helium? Does this answer agree with the lifetime you calculated in [\[link\]](#)?

Exercise:**Problem:**

Every second, the Sun converts 4 million tons of matter to energy. How long will it take the Sun to reduce its mass by 1% (the mass of the Sun is 2×10^{30} kg)? Compare your answer with the lifetime of the Sun so far.

Exercise:**Problem:**

Raymond Davis Jr.'s neutrino detector contained approximately 10^{30} chlorine atoms. During his experiment, he found that one neutrino reacted with a chlorine atom to produce one argon atom each day.

- A. How many days would he have to run the experiment for 1% of his tank to be filled with argon atoms?
- B. Convert your answer from A. into years.
- C. Compare this answer to the age of the universe, which is approximately 14 billion years (1.4×10^{10} y).
- D. What does this tell you about how frequently neutrinos interact with matter?

Glossary

helioseismology

study of pulsations or oscillations of the Sun in order to determine the characteristics of the solar interior

Work: The Scientific Definition

- Explain how an object must be displaced for a force on it to do work.
- Explain how relative directions of force and displacement determine whether the work done is positive, negative, or zero.

What It Means to Do Work

The scientific definition of work differs in some ways from its everyday meaning. Certain things we think of as hard work, such as writing an exam or carrying a heavy load on level ground, are not work as defined by a scientist. The scientific definition of work reveals its relationship to energy—whenever work is done, energy is transferred.

For work, in the scientific sense, to be done, a force must be exerted and there must be displacement in the direction of the force.

Formally, the **work** done on a system by a constant force is defined to be *the product of the component of the force in the direction of motion times the distance through which the force acts*. For one-way motion in one dimension, this is expressed in equation form as

Equation:

$$W = | \mathbf{F} | (\cos \theta) | \mathbf{d} |,$$

where W is work, $\mathbf{d}$ is the displacement of the system, and θ is the angle between the force vector $\mathbf{F}$ and the displacement vector $\mathbf{d}$, as in [\[link\]](#). We can also write this as

Equation:

$$W = Fd \cos \theta.$$

To find the work done on a system that undergoes motion that is not one-way or that is in two or three dimensions, we divide the motion into one-way one-dimensional segments and add up the work done over each segment.

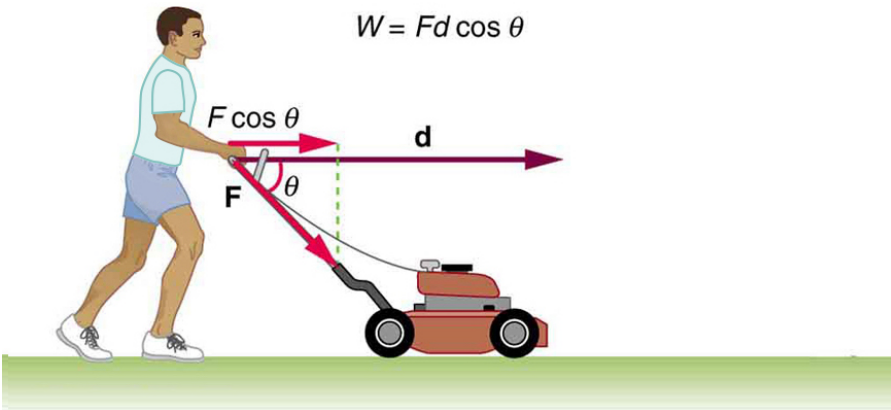
Note:**What is Work?**

The work done on a system by a constant force is *the product of the component of the force in the direction of motion times the distance through which the force acts*. For one-way motion in one dimension, this is expressed in equation form as

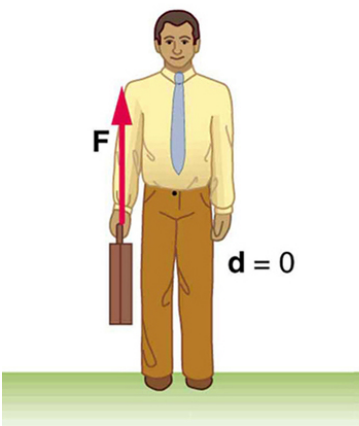
Equation:

$$W = Fd \cos \theta,$$

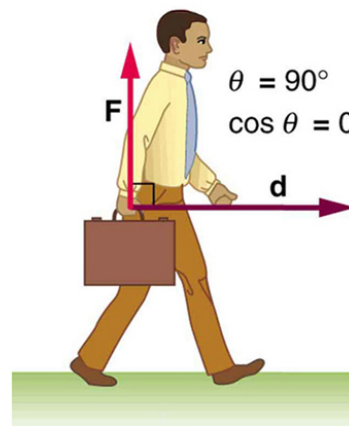
where W is work, F is the magnitude of the force on the system, d is the magnitude of the displacement of the system, and θ is the angle between the force vector $\mathbf{F}$ and the displacement vector $\mathbf{d}$.



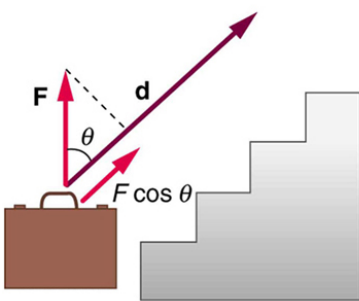
(a)



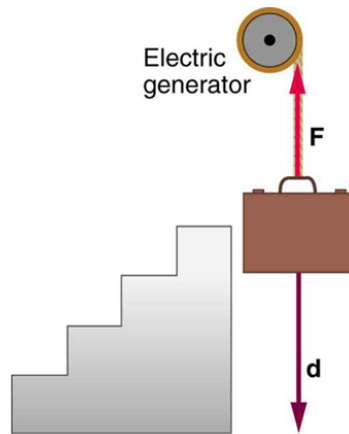
(b)



(c)



(d)



(e)

Examples of work. (a) The work done by the force $\mathbf{F}$ on this lawn mower is $Fd \cos \theta$. Note that $F \cos \theta$ is the component of the force in the direction of motion. (b) A person holding a briefcase does no work on it, because there is no

displacement. No energy is transferred to or from the briefcase. (c) The person moving the briefcase horizontally at a constant speed does no work on it, and transfers no energy to it. (d) Work is done on the briefcase by carrying it up stairs at constant speed, because there is necessarily a component of force $\mathbf{F}$ in the direction of the motion. Energy is transferred to the briefcase and could in turn be used to do work. (e) When the briefcase is lowered, energy is transferred out of the briefcase and into an electric generator. Here the work done on the briefcase by the generator is negative, removing energy from the briefcase, because $\mathbf{F}$ and $\mathbf{d}$ are in opposite directions.

To examine what the definition of work means, let us consider the other situations shown in [\[link\]](#). The person holding the briefcase in [\[link\]\(b\)](#) does no work, for example. Here $d = 0$, so $W = 0$. Why is it you get tired just holding a load? The answer is that your muscles are doing work against one another, *but they are doing no work on the system of interest* (the “briefcase-Earth system”—see [Gravitational Potential Energy](#) for more details). There must be displacement for work to be done, and there must be a component of the force in the direction of the motion. For example, the person carrying the briefcase on level ground in [\[link\]\(c\)](#) does no work on it, because the force is perpendicular to the motion. That is, $\cos 90^\circ = 0$, and so $W = 0$.

In contrast, when a force exerted on the system has a component in the direction of motion, such as in [\[link\]\(d\)](#), work is done—energy is transferred to the briefcase. Finally, in [\[link\]\(e\)](#), energy is transferred from the briefcase to a generator. There are two good ways to interpret this energy transfer. One interpretation is that the briefcase’s weight does work on the generator, giving it energy. The other interpretation is that the generator does negative work on the briefcase, thus removing energy from it. The drawing shows the latter, with the force from the generator upward

on the briefcase, and the displacement downward. This makes $\theta = 180^\circ$, and $\cos 180^\circ = -1$; therefore, W is negative.

Calculating Work

Work and energy have the same units. From the definition of work, we see that those units are force times distance. Thus, in SI units, work and energy are measured in **newton-meters**. A newton-meter is given the special name **joule** (J), and $1 \text{ J} = 1 \text{ N} \cdot \text{m} = 1 \text{ kg} \cdot \text{m}^2/\text{s}^2$. One joule is not a large amount of energy; it would lift a small 100-gram apple a distance of about 1 meter.

Example:

Calculating the Work You Do to Push a Lawn Mower Across a Large Lawn

How much work is done on the lawn mower by the person in [\[link\]](#)(a) if he exerts a constant force of 75.0 N at an angle 35° below the horizontal and pushes the mower 25.0 m on level ground? Convert the amount of work from joules to kilocalories and compare it with this person's average daily intake of 10,000 kJ (about 2400 kcal) of food energy. One *calorie* (1 cal) of heat is the amount required to warm 1 g of water by 1°C , and is equivalent to 4.186 J, while one *food calorie* (1 kcal) is equivalent to 4186 J.

Strategy

We can solve this problem by substituting the given values into the definition of work done on a system, stated in the equation $W = Fd \cos \theta$. The force, angle, and displacement are given, so that only the work W is unknown.

Solution

The equation for the work is

Equation:

$$W = Fd \cos \theta.$$

Substituting the known values gives

Equation:

$$\begin{aligned} W &= (75.0 \text{ N})(25.0 \text{ m}) \cos (35.0^\circ) \\ &= 1536 \text{ J} = 1.54 \times 10^3 \text{ J}. \end{aligned}$$

Converting the work in joules to kilocalories yields

$W = (1536 \text{ J})(1 \text{ kcal}/4186 \text{ J}) = 0.367 \text{ kcal}$. The ratio of the work done to the daily consumption is

Equation:

$$\frac{W}{2400 \text{ kcal}} = 1.53 \times 10^{-4}.$$

Discussion

This ratio is a tiny fraction of what the person consumes, but it is typical. Very little of the energy released in the consumption of food is used to do work. Even when we “work” all day long, less than 10% of our food energy intake is used to do work and more than 90% is converted to thermal energy or stored as chemical energy in fat.

Section Summary

- Work is the transfer of energy by a force acting on an object as it is displaced.
- The work W that a force $\mathbf{F}$ does on an object is the product of the magnitude F of the force, times the magnitude d of the displacement, times the cosine of the angle θ between them. In symbols,

Equation:

$$W = Fd \cos \theta.$$

- The SI unit for work and energy is the joule (J), where $1 \text{ J} = 1 \text{ N} \cdot \text{m} = 1 \text{ kg} \cdot \text{m}^2/\text{s}^2$.
- The work done by a force is zero if the displacement is either zero or perpendicular to the force.

- The work done is positive if the force and displacement have the same direction, and negative if they have opposite direction.

Conceptual Questions

Exercise:

Problem:

Give an example of something we think of as work in everyday circumstances that is not work in the scientific sense. Is energy transferred or changed in form in your example? If so, explain how this is accomplished without doing work.

Exercise:

Problem:

Give an example of a situation in which there is a force and a displacement, but the force does no work. Explain why it does no work.

Exercise:

Problem:

Describe a situation in which a force is exerted for a long time but does no work. Explain.

Problems & Exercises

Exercise:

Problem:

How much work does a supermarket checkout attendant do on a can of soup he pushes 0.600 m horizontally with a force of 5.00 N? Express your answer in joules and kilocalories.

Solution:

Equation:

$$3.00 \text{ J} = 7.17 \times 10^{-4} \text{ kcal}$$

Exercise:**Problem:**

A 75.0-kg person climbs stairs, gaining 2.50 meters in height. Find the work done to accomplish this task.

Exercise:**Problem:**

(a) Calculate the work done on a 1500-kg elevator car by its cable to lift it 40.0 m at constant speed, assuming friction averages 100 N. (b) What is the work done on the lift by the gravitational force in this process? (c) What is the total work done on the lift?

Solution:

(a) $5.92 \times 10^5 \text{ J}$

(b) $-5.88 \times 10^5 \text{ J}$

(c) The net force is zero.

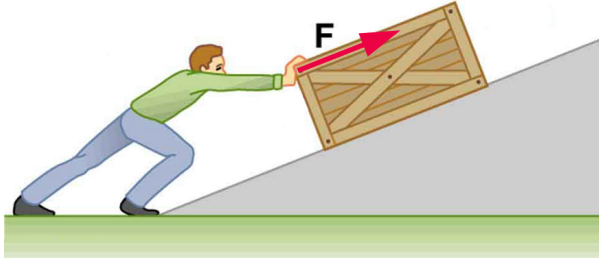
Exercise:**Problem:**

Suppose a car travels 108 km at a speed of 30.0 m/s, and uses 2.0 gal of gasoline. Only 30% of the gasoline goes into useful work by the force that keeps the car moving at constant speed despite friction. (See [\[link\]](#) for the energy content of gasoline.) (a) What is the magnitude of the force exerted to keep the car moving at constant speed? (b) If the required force is directly proportional to speed, how many gallons will be used to drive 108 km at a speed of 28.0 m/s?

Exercise:

Problem:

Calculate the work done by an 85.0-kg man who pushes a crate 4.00 m up along a ramp that makes an angle of 20.0° with the horizontal. (See [\[link\]](#).) He exerts a force of 500 N on the crate parallel to the ramp and moves at a constant speed. Be certain to include the work he does on the crate *and* on his body to get up the ramp.



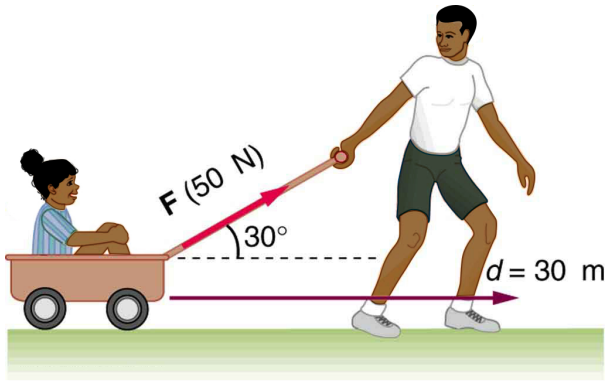
A man pushes a crate up a ramp.

Solution:**Equation:**

$$3.14 \times 10^3 \text{ J}$$

Exercise:**Problem:**

How much work is done by the boy pulling his sister 30.0 m in a wagon as shown in [\[link\]](#)? Assume no friction acts on the wagon.



The boy does work on the system of the wagon and the child when he pulls them as shown.

Exercise:

Problem:

A shopper pushes a grocery cart 20.0 m at constant speed on level ground, against a 35.0 N frictional force. He pushes in a direction 25.0° below the horizontal. (a) What is the work done on the cart by friction? (b) What is the work done on the cart by the gravitational force? (c) What is the work done on the cart by the shopper? (d) Find the force the shopper exerts, using energy considerations. (e) What is the total work done on the cart?

Solution:

(a) -700 J

(b) 0

(c) 700 J

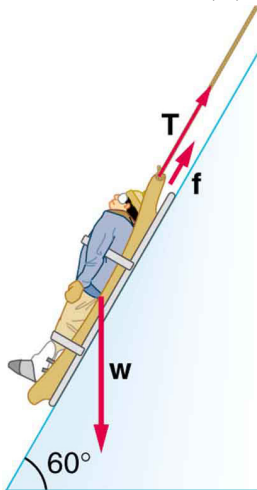
(d) 38.6 N

(e) 0

Exercise:

Problem:

Suppose the ski patrol lowers a rescue sled and victim, having a total mass of 90.0 kg , down a 60.0° slope at constant speed, as shown in [\[link\]](#). The coefficient of friction between the sled and the snow is 0.100 . (a) How much work is done by friction as the sled moves 30.0 m along the hill? (b) How much work is done by the rope on the sled in this distance? (c) What is the work done by the gravitational force on the sled? (d) What is the total work done?



A rescue sled and victim are lowered down a steep slope.

Glossary

energy

the ability to do work

work

the transfer of energy by a force that causes an object to be displaced;
the product of the component of the force in the direction of the
displacement and the magnitude of the displacement

joule

SI unit of work and energy, equal to one newton-meter

Kinetic Energy and the Work-Energy Theorem

- Explain work as a transfer of energy and net work as the work done by the net force.
- Explain and apply the work-energy theorem.

Work Transfers Energy

What happens to the work done on a system? Energy is transferred into the system, but in what form? Does it remain in the system or move on? The answers depend on the situation. For example, if the lawn mower in [\[link\]](#) (a) is pushed just hard enough to keep it going at a constant speed, then energy put into the mower by the person is removed continuously by friction, and eventually leaves the system in the form of heat transfer. In contrast, work done on the briefcase by the person carrying it up stairs in [\[link\]](#) (d) is stored in the briefcase-Earth system and can be recovered at any time, as shown in [\[link\]](#) (e). In fact, the building of the pyramids in ancient Egypt is an example of storing energy in a system by doing work on the system. Some of the energy imparted to the stone blocks in lifting them during construction of the pyramids remains in the stone-Earth system and has the potential to do work.

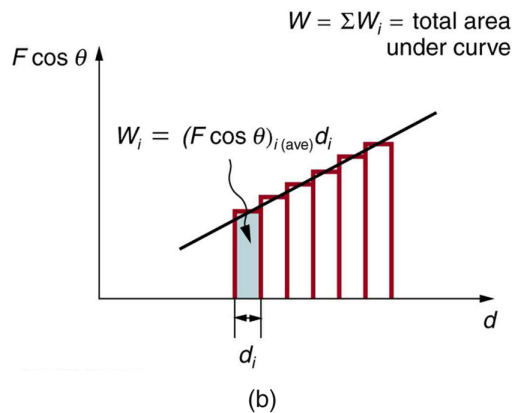
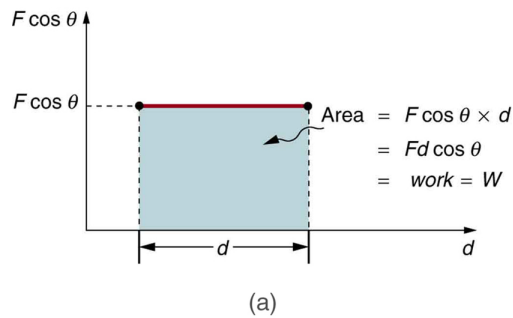
In this section we begin the study of various types of work and forms of energy. We will find that some types of work leave the energy of a system constant, for example, whereas others change the system in some way, such as making it move. We will also develop definitions of important forms of energy, such as the energy of motion.

Net Work and the Work-Energy Theorem

We know from the study of Newton's laws in [Dynamics: Force and Newton's Laws of Motion](#) that net force causes acceleration. We will see in this section that work done by the net force gives a system energy of motion, and in the process we will also find an expression for the energy of motion.

Let us start by considering the total, or net, work done on a system. Net work is defined to be the sum of work done by all external forces—that is, **net work** is the work done by the net external force $\mathbf{F}_{\text{net}}$. In equation form, this is $W_{\text{net}} = F_{\text{net}}d \cos \theta$ where θ is the angle between the force vector and the displacement vector.

[\[link\]](#)(a) shows a graph of force versus displacement for the component of the force in the direction of the displacement—that is, an $F \cos \theta$ vs. d graph. In this case, $F \cos \theta$ is constant. You can see that the area under the graph is $Fd \cos \theta$, or the work done. [\[link\]](#)(b) shows a more general process where the force varies. The area under the curve is divided into strips, each having an average force $(F \cos \theta)_{i(\text{ave})}$. The work done is $(F \cos \theta)_{i(\text{ave})}d_i$ for each strip, and the total work done is the sum of the W_i . Thus the total work done is the total area under the curve, a useful property to which we shall refer later.

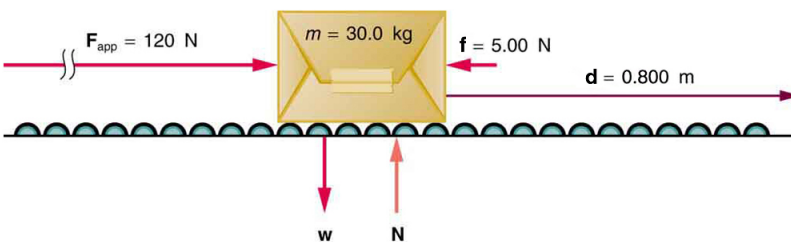


(a) A graph of $F \cos \theta$ vs. d , when $F \cos \theta$ is

constant. The area under the curve represents the work done by the force.

(b) A graph of $F \cos \theta$ vs. d in which the force varies. The work done for each interval is the area of each strip; thus, the total area under the curve equals the total work done.

Net work will be simpler to examine if we consider a one-dimensional situation where a force is used to accelerate an object in a direction parallel to its initial velocity. Such a situation occurs for the package on the roller belt conveyor system shown in [\[link\]](#).



A package on a roller belt is pushed horizontally through a distance d .

The force of gravity and the normal force acting on the package are perpendicular to the displacement and do no work. Moreover, they are also equal in magnitude and opposite in direction so they cancel in calculating the net force. The net force arises solely from the horizontal applied force F_{app} and the horizontal friction force f . Thus, as expected, the net force is

parallel to the displacement, so that $\theta = 0^\circ$ and $\cos \theta = 1$, and the net work is given by

Equation:

$$W_{\text{net}} = F_{\text{net}}d.$$

The effect of the net force $\mathbf{F}_{\text{net}}$ is to accelerate the package from v_0 to v . The kinetic energy of the package increases, indicating that the net work done on the system is positive. (See [\[link\]](#).) By using Newton's second law, and doing some algebra, we can reach an interesting conclusion. Substituting $F_{\text{net}} = ma$ from Newton's second law gives

Equation:

$$W_{\text{net}} = mad.$$

To get a relationship between net work and the speed given to a system by the net force acting on it, we take $d = x - x_0$ and use the equation studied in [Motion Equations for Constant Acceleration in One Dimension](#) for the change in speed over a distance d if the acceleration has the constant value a ; namely, $v^2 = v_0^2 + 2ad$ (note that a appears in the expression for the net work). Solving for acceleration gives $a = \frac{v^2 - v_0^2}{2d}$. When a is substituted into the preceding expression for W_{net} , we obtain

Equation:

$$W_{\text{net}} = m \left(\frac{v^2 - v_0^2}{2d} \right) d.$$

The d cancels, and we rearrange this to obtain

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2.$$

This expression is called the **work-energy theorem**, and it actually applies *in general* (even for forces that vary in direction and magnitude), although we have derived it for the special case of a constant force parallel to the displacement. The theorem implies that the net work on a system equals the change in the quantity $\frac{1}{2}mv^2$. This quantity is our first example of a form of energy.

Note:

The Work-Energy Theorem

The net work on a system equals the change in the quantity $\frac{1}{2}mv^2$.

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$$

The quantity $\frac{1}{2}mv^2$ in the work-energy theorem is defined to be the translational **kinetic energy** (KE) of a mass m moving at a speed v . (*Translational* kinetic energy is distinct from *rotational* kinetic energy, which is considered later.) In equation form, the translational kinetic energy,

Equation:

$$\text{KE} = \frac{1}{2}mv^2,$$

is the energy associated with translational motion. Kinetic energy is a form of energy associated with the motion of a particle, single body, or system of objects moving together.

We are aware that it takes energy to get an object, like a car or the package in [\[link\]](#), up to speed, but it may be a bit surprising that kinetic energy is proportional to speed squared. This proportionality means, for example, that a car traveling at 100 km/h has four times the kinetic energy it has at 50

km/h, helping to explain why high-speed collisions are so devastating. We will now consider a series of examples to illustrate various aspects of work and energy.

Example:

Calculating the Kinetic Energy of a Package

Suppose a 30.0-kg package on the roller belt conveyor system in [\[link\]](#) is moving at 0.500 m/s. What is its kinetic energy?

Strategy

Because the mass m and speed v are given, the kinetic energy can be calculated from its definition as given in the equation $\text{KE} = \frac{1}{2}mv^2$.

Solution

The kinetic energy is given by

Equation:

$$\text{KE} = \frac{1}{2}mv^2.$$

Entering known values gives

Equation:

$$\text{KE} = 0.5(30.0 \text{ kg})(0.500 \text{ m/s})^2,$$

which yields

Equation:

$$\text{KE} = 3.75 \text{ kg} \cdot \text{m}^2/\text{s}^2 = 3.75 \text{ J}.$$

Discussion

Note that the unit of kinetic energy is the joule, the same as the unit of work, as mentioned when work was first defined. It is also interesting that, although this is a fairly massive package, its kinetic energy is not large at this relatively low speed. This fact is consistent with the observation that people can move packages like this without exhausting themselves.

Example:**Determining the Work to Accelerate a Package**

Suppose that you push on the 30.0-kg package in [\[link\]](#) with a constant force of 120 N through a distance of 0.800 m, and that the opposing friction force averages 5.00 N.

(a) Calculate the net work done on the package. (b) Solve the same problem as in part (a), this time by finding the work done by each force that contributes to the net force.

Strategy and Concept for (a)

This is a motion in one dimension problem, because the downward force (from the weight of the package) and the normal force have equal magnitude and opposite direction, so that they cancel in calculating the net force, while the applied force, friction, and the displacement are all horizontal. (See [\[link\]](#).) As expected, the net work is the net force times distance.

Solution for (a)

The net force is the push force minus friction, or

$F_{\text{net}} = 120 \text{ N} - 5.00 \text{ N} = 115 \text{ N}$. Thus the net work is

Equation:

$$\begin{aligned} W_{\text{net}} &= F_{\text{net}}d = (115 \text{ N})(0.800 \text{ m}) \\ &= 92.0 \text{ N} \cdot \text{m} = 92.0 \text{ J}. \end{aligned}$$

Discussion for (a)

This value is the net work done on the package. The person actually does more work than this, because friction opposes the motion. Friction does negative work and removes some of the energy the person expends and converts it to thermal energy. The net work equals the sum of the work done by each individual force.

Strategy and Concept for (b)

The forces acting on the package are gravity, the normal force, the force of friction, and the applied force. The normal force and force of gravity are each perpendicular to the displacement, and therefore do no work.

Solution for (b)

The applied force does work.

Equation:

$$\begin{aligned}
 W_{\text{app}} &= F_{\text{app}} d \cos(0^\circ) = F_{\text{app}} d \\
 &= (120 \text{ N})(0.800 \text{ m}) \\
 &= 96.0 \text{ J}
 \end{aligned}$$

The friction force and displacement are in opposite directions, so that $\theta = 180^\circ$, and the work done by friction is

Equation:

$$\begin{aligned}
 W_{\text{fr}} &= F_{\text{fr}} d \cos(180^\circ) = -F_{\text{fr}} d \\
 &= -(5.00 \text{ N})(0.800 \text{ m}) \\
 &= -4.00 \text{ J.}
 \end{aligned}$$

So the amounts of work done by gravity, by the normal force, by the applied force, and by friction are, respectively,

Equation:

$$\begin{aligned}
 W_{\text{gr}} &= 0, \\
 W_{\text{N}} &= 0, \\
 W_{\text{app}} &= 96.0 \text{ J}, \\
 W_{\text{fr}} &= -4.00 \text{ J.}
 \end{aligned}$$

The total work done as the sum of the work done by each force is then seen to be

Equation:

$$W_{\text{total}} = W_{\text{gr}} + W_{\text{N}} + W_{\text{app}} + W_{\text{fr}} = 92.0 \text{ J.}$$

Discussion for (b)

The calculated total work W_{total} as the sum of the work by each force agrees, as expected, with the work W_{net} done by the net force. The work done by a collection of forces acting on an object can be calculated by either approach.

Example:

Determining Speed from Work and Energy

Find the speed of the package in [\[link\]](#) at the end of the push, using work and energy concepts.

Strategy

Here the work-energy theorem can be used, because we have just calculated the net work, W_{net} , and the initial kinetic energy, $\frac{1}{2}mv_0^2$. These calculations allow us to find the final kinetic energy, $\frac{1}{2}mv^2$, and thus the final speed v .

Solution

The work-energy theorem in equation form is

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2.$$

Solving for $\frac{1}{2}mv^2$ gives

Equation:

$$\frac{1}{2}mv^2 = W_{\text{net}} + \frac{1}{2}mv_0^2.$$

Thus,

Equation:

$$\frac{1}{2}mv^2 = 92.0 \text{ J} + 3.75 \text{ J} = 95.75 \text{ J}.$$

Solving for the final speed as requested and entering known values gives

Equation:

$$\begin{aligned} v &= \sqrt{\frac{2(95.75 \text{ J})}{m}} = \sqrt{\frac{191.5 \text{ kg}\cdot\text{m}^2/\text{s}^2}{30.0 \text{ kg}}} \\ &= 2.53 \text{ m/s}. \end{aligned}$$

Discussion

Using work and energy, we not only arrive at an answer, we see that the final kinetic energy is the sum of the initial kinetic energy and the net work

done on the package. This means that the work indeed adds to the energy of the package.

Example:

Work and Energy Can Reveal Distance, Too

How far does the package in [\[link\]](#) coast after the push, assuming friction remains constant? Use work and energy considerations.

Strategy

We know that once the person stops pushing, friction will bring the package to rest. In terms of energy, friction does negative work until it has removed all of the package's kinetic energy. The work done by friction is the force of friction times the distance traveled times the cosine of the angle between the friction force and displacement; hence, this gives us a way of finding the distance traveled after the person stops pushing.

Solution

The normal force and force of gravity cancel in calculating the net force. The horizontal friction force is then the net force, and it acts opposite to the displacement, so $\theta = 180^\circ$. To reduce the kinetic energy of the package to zero, the work W_{fr} by friction must be minus the kinetic energy that the package started with plus what the package accumulated due to the pushing. Thus $W_{\text{fr}} = -95.75 \text{ J}$. Furthermore, $W_{\text{fr}} = f d' \cos \theta = -f d'$, where d' is the distance it takes to stop. Thus,

Equation:

$$d' = -\frac{W_{\text{fr}}}{f} = -\frac{-95.75 \text{ J}}{5.00 \text{ N}},$$

and so

Equation:

$$d' = 19.2 \text{ m}.$$

Discussion

This is a reasonable distance for a package to coast on a relatively friction-free conveyor system. Note that the work done by friction is negative (the

force is in the opposite direction of motion), so it removes the kinetic energy.

Some of the examples in this section can be solved without considering energy, but at the expense of missing out on gaining insights about what work and energy are doing in this situation. On the whole, solutions involving energy are generally shorter and easier than those using kinematics and dynamics alone.

Section Summary

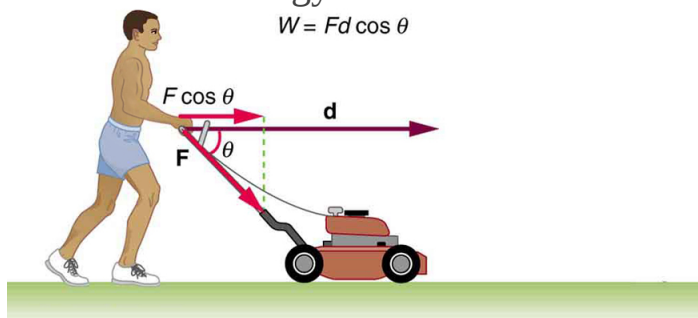
- The net work W_{net} is the work done by the net force acting on an object.
- Work done on an object transfers energy to the object.
- The translational kinetic energy of an object of mass m moving at speed v is $\text{KE} = \frac{1}{2}mv^2$.
- The work-energy theorem states that the net work W_{net} on a system changes its kinetic energy, $W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$.

Conceptual Questions

Exercise:

Problem:

The person in [\[link\]](#) does work on the lawn mower. Under what conditions would the mower gain energy? Under what conditions would it lose energy?



Exercise:**Problem:**

Work done on a system puts energy into it. Work done by a system removes energy from it. Give an example for each statement.

Exercise:**Problem:**

When solving for speed in [\[link\]](#), we kept only the positive root. Why?

Problems & Exercises**Exercise:****Problem:**

Compare the kinetic energy of a 20,000-kg truck moving at 110 km/h with that of an 80.0-kg astronaut in orbit moving at 27,500 km/h.

Solution:

1/250

Exercise:**Problem:**

(a) How fast must a 3000-kg elephant move to have the same kinetic energy as a 65.0-kg sprinter running at 10.0 m/s? (b) Discuss how the larger energies needed for the movement of larger animals would relate to metabolic rates.

Exercise:**Problem:**

Confirm the value given for the kinetic energy of an aircraft carrier in [\[link\]](#). You will need to look up the definition of a nautical mile (1 knot = 1 nautical mile/h).

Solution:

$$1.1 \times 10^{10} \text{ J}$$

Exercise:**Problem:**

(a) Calculate the force needed to bring a 950-kg car to rest from a speed of 90.0 km/h in a distance of 120 m (a fairly typical distance for a non-panic stop). (b) Suppose instead the car hits a concrete abutment at full speed and is brought to a stop in 2.00 m. Calculate the force exerted on the car and compare it with the force found in part (a).

Exercise:**Problem:**

A car's bumper is designed to withstand a 4.0-km/h (1.12-m/s) collision with an immovable object without damage to the body of the car. The bumper cushions the shock by absorbing the force over a distance. Calculate the magnitude of the average force on a bumper that collapses 0.200 m while bringing a 900-kg car to rest from an initial speed of 1.12 m/s.

Solution:

$$2.8 \times 10^3 \text{ N}$$

Exercise:

Problem:

Boxing gloves are padded to lessen the force of a blow. (a) Calculate the force exerted by a boxing glove on an opponent's face, if the glove and face compress 7.50 cm during a blow in which the 7.00-kg arm and glove are brought to rest from an initial speed of 10.0 m/s. (b) Calculate the force exerted by an identical blow in the days when no gloves were used and the knuckles and face would compress only 2.00 cm. (c) Discuss the magnitude of the force with glove on. Does it seem high enough to cause damage even though it is lower than the force with no glove?

Exercise:**Problem:**

Using energy considerations, calculate the average force a 60.0-kg sprinter exerts backward on the track to accelerate from 2.00 to 8.00 m/s in a distance of 25.0 m, if he encounters a headwind that exerts an average force of 30.0 N against him.

Solution:

102 N

Glossary

net work

work done by the net force, or vector sum of all the forces, acting on an object

work-energy theorem

the result, based on Newton's laws, that the net work done on an object is equal to its change in kinetic energy

kinetic energy

the energy an object has by reason of its motion, equal to $\frac{1}{2}mv^2$ for the translational (i.e., non-rotational) motion of an object of mass m moving at speed v

Gravitational Potential Energy

- Explain gravitational potential energy in terms of work done against gravity.
- Show that the gravitational potential energy of an object of mass m at height h on Earth is given by $PE_g = mgh$.
- Show how knowledge of the potential energy as a function of position can be used to simplify calculations and explain physical phenomena.

Work Done Against Gravity

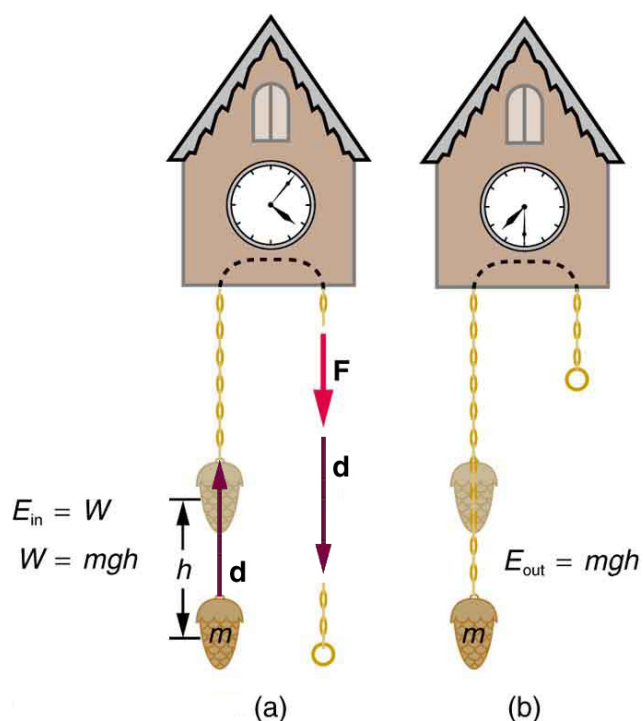
Climbing stairs and lifting objects is work in both the scientific and everyday sense—it is work done against the gravitational force. When there is work, there is a transformation of energy. The work done against the gravitational force goes into an important form of stored energy that we will explore in this section.

Let us calculate the work done in lifting an object of mass m through a height h , such as in [\[link\]](#). If the object is lifted straight up at constant speed, then the force needed to lift it is equal to its weight mg . The work done on the mass is then $W = Fd = mgh$. We define this to be the **gravitational potential energy** (PE_g) put into (or gained by) the object-Earth system. This energy is associated with the state of separation between two objects that attract each other by the gravitational force. For convenience, we refer to this as the PE_g gained by the object, recognizing that this is energy stored in the gravitational field of Earth. Why do we use the word “system”? Potential energy is a property of a system rather than of a single object—due to its physical position. An object’s gravitational potential is due to its position relative to the surroundings within the Earth-object system. The force applied to the object is an external force, from outside the system. When it does positive work it increases the gravitational potential energy of the system. Because gravitational potential energy depends on relative position, we need a reference level at which to set the potential energy equal to 0. We usually choose this point to be Earth’s surface, but this point is arbitrary; what is important is the *difference* in gravitational potential energy, because this difference is what relates to the work done. The difference in gravitational potential energy of an object (in the Earth-object system) between two rungs of a ladder will be the same for the first two rungs as for the last two rungs.

Converting Between Potential Energy and Kinetic Energy

Gravitational potential energy may be converted to other forms of energy, such as kinetic energy. If we release the mass, gravitational force will do an amount of work

equal to mgh on it, thereby increasing its kinetic energy by that same amount (by the work-energy theorem). We will find it more useful to consider just the conversion of PE_g to KE without explicitly considering the intermediate step of work. (See [\[link\]](#).) This shortcut makes it is easier to solve problems using energy (if possible) rather than explicitly using forces.



(a) The work done to lift the weight is stored in the mass-Earth system as gravitational potential energy. (b) As the weight moves downward, this gravitational potential energy is transferred to the cuckoo clock.

More precisely, we define the *change* in gravitational potential energy ΔPE_g to be
Equation:

$$\Delta PE_g = mgh,$$

where, for simplicity, we denote the change in height by h rather than the usual Δh . Note that h is positive when the final height is greater than the initial height, and vice versa. For example, if a 0.500-kg mass hung from a cuckoo clock is raised 1.00 m, then its change in gravitational potential energy is

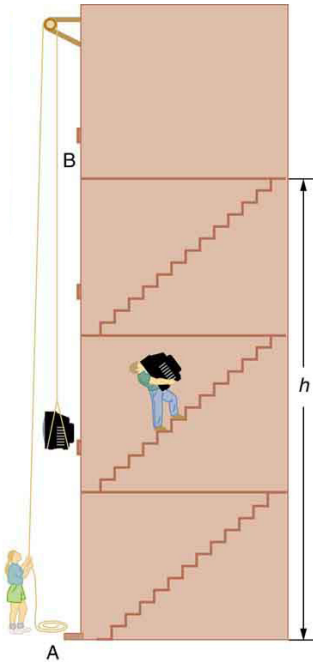
Equation:

$$\begin{aligned} mgh &= (0.500 \text{ kg}) (9.80 \text{ m/s}^2) (1.00 \text{ m}) \\ &= 4.90 \text{ kg} \cdot \text{m}^2/\text{s}^2 = 4.90 \text{ J}. \end{aligned}$$

Note that the units of gravitational potential energy turn out to be joules, the same as for work and other forms of energy. As the clock runs, the mass is lowered. We can think of the mass as gradually giving up its 4.90 J of gravitational potential energy, *without directly considering the force of gravity that does the work*.

Using Potential Energy to Simplify Calculations

The equation $\Delta \text{PE}_g = mgh$ applies for any path that has a change in height of h , not just when the mass is lifted straight up. (See [\[link\]](#).) It is much easier to calculate mgh (a simple multiplication) than it is to calculate the work done along a complicated path. The idea of gravitational potential energy has the double advantage that it is very broadly applicable and it makes calculations easier. From now on, we will consider that any change in vertical position h of a mass m is accompanied by a change in gravitational potential energy mgh , and we will avoid the equivalent but more difficult task of calculating work done by or against the gravitational force.



The change in
gravitational
potential energy
(ΔPE_g)
between points
A and B is
independent of
the path.

$\Delta PE_g = mgh$
for any path
between the two
points. Gravity
is one of a small
class of forces
where the work
done by or
against the force
depends only on
the starting and
ending points,
not on the path
between them.

Example:**The Force to Stop Falling**

A 60.0-kg person jumps onto the floor from a height of 3.00 m. If he lands stiffly (with his knee joints compressing by 0.500 cm), calculate the force on the knee joints.

Strategy

This person's energy is brought to zero in this situation by the work done on him by the floor as he stops. The initial PE_g is transformed into KE as he falls. The work done by the floor reduces this kinetic energy to zero.

Solution

The work done on the person by the floor as he stops is given by

Equation:

$$W = Fd \cos \theta = -Fd,$$

with a minus sign because the displacement while stopping and the force from floor are in opposite directions ($\cos \theta = \cos 180^\circ = -1$). The floor removes energy from the system, so it does negative work.

The kinetic energy the person has upon reaching the floor is the amount of potential energy lost by falling through height h :

Equation:

$$KE = -\Delta PE_g = -mgh,$$

The distance d that the person's knees bend is much smaller than the height h of the fall, so the additional change in gravitational potential energy during the knee bend is ignored.

The work W done by the floor on the person stops the person and brings the person's kinetic energy to zero:

Equation:

$$W = -KE = mgh.$$

Combining this equation with the expression for W gives

Equation:

$$-Fd = mgh.$$

Recalling that h is negative because the person fell *down*, the force on the knee joints is given by

Equation:

$$F = -\frac{mgh}{d} = -\frac{(60.0 \text{ kg})(9.80 \text{ m/s}^2)(-3.00 \text{ m})}{5.00 \times 10^{-3} \text{ m}} = 3.53 \times 10^5 \text{ N}.$$

Discussion

Such a large force (500 times more than the person's weight) over the short impact time is enough to break bones. A much better way to cushion the shock is by bending the legs or rolling on the ground, increasing the time over which the force acts. A bending motion of 0.5 m this way yields a force 100 times smaller than in the example. A kangaroo's hopping shows this method in action. The kangaroo is the only large animal to use hopping for locomotion, but the shock in hopping is cushioned by the bending of its hind legs in each jump. (See [\[link\]](#).)

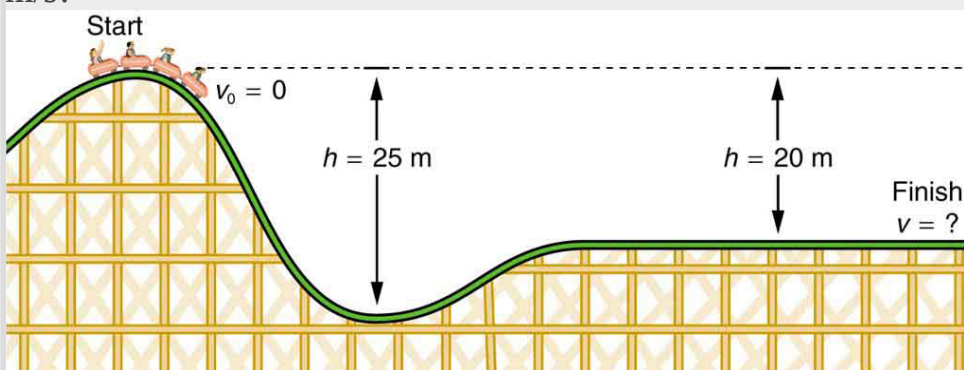


The work done by the ground upon the kangaroo reduces its kinetic energy to zero as it lands. However, by applying the force of the ground on the hind legs over a longer distance, the impact on the bones is reduced.
(credit: Chris Samuel, Flickr)

Example:

Finding the Speed of a Roller Coaster from its Height

(a) What is the final speed of the roller coaster shown in [\[link\]](#) if it starts from rest at the top of the 20.0 m hill and work done by frictional forces is negligible? (b) What is its final speed (again assuming negligible friction) if its initial speed is 5.00 m/s?



The speed of a roller coaster increases as gravity pulls it downhill and is greatest at its lowest point. Viewed in terms of energy, the roller-coaster-Earth system's gravitational potential energy is converted to kinetic energy. If work done by friction is negligible, all ΔPE_g is converted to KE.

Strategy

The roller coaster loses potential energy as it goes downhill. We neglect friction, so that the remaining force exerted by the track is the normal force, which is perpendicular to the direction of motion and does no work. The net work on the roller coaster is then done by gravity alone. The *loss* of gravitational potential energy from moving *downward* through a distance h equals the *gain* in kinetic energy. This can be written in equation form as $-\Delta PE_g = \Delta KE$. Using the equations for PE_g and KE, we can solve for the final speed v , which is the desired quantity.

Solution for (a)

Here the initial kinetic energy is zero, so that $\Delta KE = \frac{1}{2}mv^2$. The equation for change in potential energy states that $\Delta PE_g = mgh$. Since h is negative in this case, we will rewrite this as $\Delta PE_g = -mg |h|$ to show the minus sign clearly. Thus,

Equation:

$$-\Delta\text{PE}_g = \Delta\text{KE}$$

becomes

Equation:

$$mg | h | = \frac{1}{2}mv^2.$$

Solving for v , we find that mass cancels and that

Equation:

$$v = \sqrt{2g | h |}.$$

Substituting known values,

Equation:

$$\begin{aligned} v &= \sqrt{2(9.80 \text{ m/s}^2)(20.0 \text{ m})} \\ &= 19.8 \text{ m/s.} \end{aligned}$$

Solution for (b)

Again $-\Delta\text{PE}_g = \Delta\text{KE}$. In this case there is initial kinetic energy, so

$\Delta\text{KE} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2$. Thus,

Equation:

$$mg | h | = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2.$$

Rearranging gives

Equation:

$$\frac{1}{2}mv^2 = mg | h | + \frac{1}{2}mv_0^2.$$

This means that the final kinetic energy is the sum of the initial kinetic energy and the gravitational potential energy. Mass again cancels, and

Equation:

$$v = \sqrt{2g | h | + v_0^2}.$$

This equation is very similar to the kinematics equation $v = \sqrt{v_0^2 + 2ad}$, but it is more general—the kinematics equation is valid only for constant acceleration, whereas our equation above is valid for any path regardless of whether the object moves with a constant acceleration. Now, substituting known values gives

Equation:

$$\begin{aligned} v &= \sqrt{2(9.80 \text{ m/s}^2)(20.0 \text{ m}) + (5.00 \text{ m/s})^2} \\ &= 20.4 \text{ m/s.} \end{aligned}$$

Discussion and Implications

First, note that mass cancels. This is quite consistent with observations made in [Falling Objects](#) that all objects fall at the same rate if friction is negligible. Second, only the speed of the roller coaster is considered; there is no information about its direction at any point. This reveals another general truth. When friction is negligible, the speed of a falling body depends only on its initial speed and height, and not on its mass or the path taken. For example, the roller coaster will have the same final speed whether it falls 20.0 m straight down or takes a more complicated path like the one in the figure. Third, and perhaps unexpectedly, the final speed in part (b) is greater than in part (a), but by far less than 5.00 m/s. Finally, note that speed can be found at *any* height along the way by simply using the appropriate value of h at the point of interest.

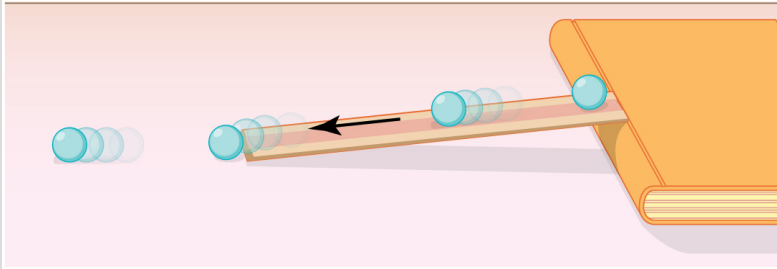
We have seen that work done by or against the gravitational force depends only on the starting and ending points, and not on the path between, allowing us to define the simplifying concept of gravitational potential energy. We can do the same thing for a few other forces, and we will see that this leads to a formal definition of the law of conservation of energy.

Note:

Making Connections: Take-Home Investigation—Converting Potential to Kinetic Energy

One can study the conversion of gravitational potential energy into kinetic energy in this experiment. On a smooth, level surface, use a ruler of the kind that has a groove running along its length and a book to make an incline (see [link](#)). Place a marble at the 10-cm position on the ruler and let it roll down the ruler. When it hits the level surface, measure the time it takes to roll one meter. Now place the marble

at the 20-cm and the 30-cm positions and again measure the times it takes to roll 1 m on the level surface. Find the velocity of the marble on the level surface for all three positions. Plot velocity squared versus the distance traveled by the marble. What is the shape of each plot? If the shape is a straight line, the plot shows that the marble's kinetic energy at the bottom is proportional to its potential energy at the release point.



A marble rolls down a ruler, and its speed on the level surface is measured.

Section Summary

- Work done against gravity in lifting an object becomes potential energy of the object-Earth system.
- The change in gravitational potential energy, ΔPE_g , is $\Delta PE_g = mgh$, with h being the increase in height and g the acceleration due to gravity.
- The gravitational potential energy of an object near Earth's surface is due to its position in the mass-Earth system. Only differences in gravitational potential energy, ΔPE_g , have physical significance.
- As an object descends without friction, its gravitational potential energy changes into kinetic energy corresponding to increasing speed, so that $\Delta KE = -\Delta PE_g$.

Conceptual Questions

Exercise:

Problem:

In [\[link\]](#), we calculated the final speed of a roller coaster that descended 20 m in height and had an initial speed of 5 m/s downhill. Suppose the roller coaster had had an initial speed of 5 m/s *uphill* instead, and it coasted uphill, stopped, and then rolled back down to a final point 20 m below the start. We would find in that case that its final speed is the same as its initial speed. Explain in terms of conservation of energy.

Exercise:**Problem:**

Does the work you do on a book when you lift it onto a shelf depend on the path taken? On the time taken? On the height of the shelf? On the mass of the book?

Problems & Exercises**Exercise:****Problem:**

A hydroelectric power facility (see [\[link\]](#)) converts the gravitational potential energy of water behind a dam to electric energy. (a) What is the gravitational potential energy relative to the generators of a lake of volume 50.0 km^3 (mass = $5.00 \times 10^{13} \text{ kg}$), given that the lake has an average height of 40.0 m above the generators? (b) Compare this with the energy stored in a 9-megaton fusion bomb.



Hydroelectric facility (credit: Denis

Solution:

(a) $1.96 \times 10^{16} \text{ J}$

(b) The ratio of gravitational potential energy in the lake to the energy stored in the bomb is 0.52. That is, the energy stored in the lake is approximately half that in a 9-megaton fusion bomb.

Exercise:

Problem:

(a) How much gravitational potential energy (relative to the ground on which it is built) is stored in the Great Pyramid of Cheops, given that its mass is about $7 \times 10^9 \text{ kg}$ and its center of mass is 36.5 m above the surrounding ground? (b) How does this energy compare with the daily food intake of a person?

Exercise:

Problem:

Suppose a 350-g kookaburra (a large kingfisher bird) picks up a 75-g snake and raises it 2.5 m from the ground to a branch. (a) How much work did the bird do on the snake? (b) How much work did it do to raise its own center of mass to the branch?

Solution:

(a) 1.8 J

(b) 8.6 J

Exercise:

Problem:

In [\[link\]](#), we found that the speed of a roller coaster that had descended 20.0 m was only slightly greater when it had an initial speed of 5.00 m/s than when it started from rest. This implies that $\Delta PE \gg KE_i$. Confirm this statement by taking the ratio of ΔPE to KE_i . (Note that mass cancels.)

Exercise:

Problem:

A 100-g toy car is propelled by a compressed spring that starts it moving. The car follows the curved track in [\[link\]](#). Show that the final speed of the toy car is 0.687 m/s if its initial speed is 2.00 m/s and it coasts up the frictionless slope, gaining 0.180 m in altitude.



A toy car moves up a sloped track.
(credit: Leszek Leszczynski, Flickr)

Solution:**Equation:**

$$v_f = \sqrt{2gh + v_0^2} = \sqrt{2(9.80 \text{ m/s}^2)(-0.180 \text{ m}) + (2.00 \text{ m/s})^2} = 0.687 \text{ m/s}$$

Exercise:**Problem:**

In a downhill ski race, surprisingly, little advantage is gained by getting a running start. (This is because the initial kinetic energy is small compared with the gain in gravitational potential energy on even small hills.) To demonstrate this, find the final speed and the time taken for a skier who skies 70.0 m along a 30° slope neglecting friction: (a) Starting from rest. (b) Starting with an initial speed of 2.50 m/s. (c) Does the answer surprise you? Discuss why it is still advantageous to get a running start in very competitive events.

Glossary

gravitational potential energy

the energy an object has due to its position in a gravitational field

Conservative Forces and Potential Energy

- Define conservative force, potential energy, and mechanical energy.
- Explain the potential energy of a spring in terms of its compression when Hooke's law applies.
- Use the work-energy theorem to show how having only conservative forces implies conservation of mechanical energy.

Potential Energy and Conservative Forces

Work is done by a force, and some forces, such as weight, have special characteristics. A **conservative force** is one, like the gravitational force, for which work done by or against it depends only on the starting and ending points of a motion and not on the path taken. We can define a **potential energy** (PE) for any conservative force, just as we did for the gravitational force. For example, when you wind up a toy, an egg timer, or an old-fashioned watch, you do work against its spring and store energy in it. (We treat these springs as ideal, in that we assume there is no friction and no production of thermal energy.) This stored energy is recoverable as work, and it is useful to think of it as potential energy contained in the spring. Indeed, the reason that the spring has this characteristic is that its force is *conservative*. That is, a conservative force results in stored or potential energy. Gravitational potential energy is one example, as is the energy stored in a spring. We will also see how conservative forces are related to the conservation of energy.

Note:

Potential Energy and Conservative Forces

Potential energy is the energy a system has due to position, shape, or configuration. It is stored energy that is completely recoverable.

A conservative force is one for which work done by or against it depends only on the starting and ending points of a motion and not on the path taken.

We can define a potential energy (PE) for any conservative force. The work done against a conservative force to reach a final configuration

depends on the configuration, not the path followed, and is the potential energy added.

Potential Energy of a Spring

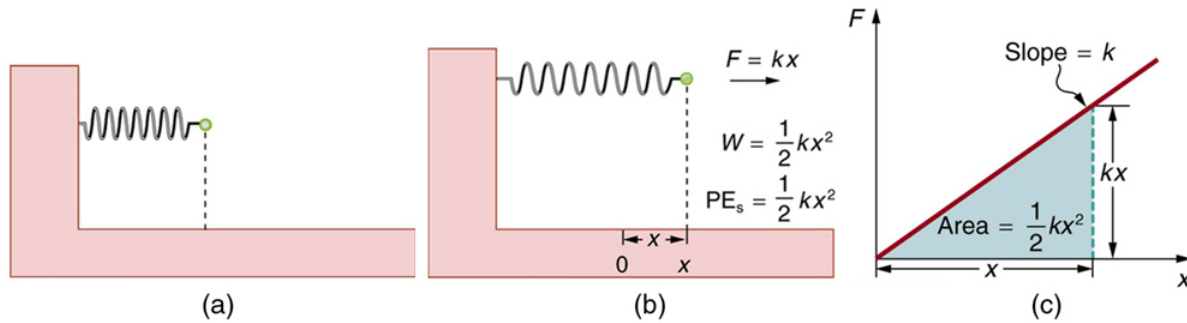
First, let us obtain an expression for the potential energy stored in a spring (PE_s). We calculate the work done to stretch or compress a spring that obeys Hooke's law. (Hooke's law was examined in [Elasticity: Stress and Strain](#), and states that the magnitude of force F on the spring and the resulting deformation ΔL are proportional, $F = k\Delta L$.) (See [\[link\]](#).) For our spring, we will replace ΔL (the amount of deformation produced by a force F) by the distance x that the spring is stretched or compressed along its length. So the force needed to stretch the spring has magnitude $F = kx$, where k is the spring's force constant. The force increases linearly from 0 at the start to kx in the fully stretched position. The average force is $kx/2$. Thus the work done in stretching or compressing the spring is

$W_s = Fd = \left(\frac{kx}{2}\right)x = \frac{1}{2}kx^2$. Alternatively, we noted in [Kinetic Energy and the Work-Energy Theorem](#) that the area under a graph of F vs. x is the work done by the force. In [\[link\]](#)(c) we see that this area is also $\frac{1}{2}kx^2$. We therefore define the **potential energy of a spring**, PE_s , to be

Equation:

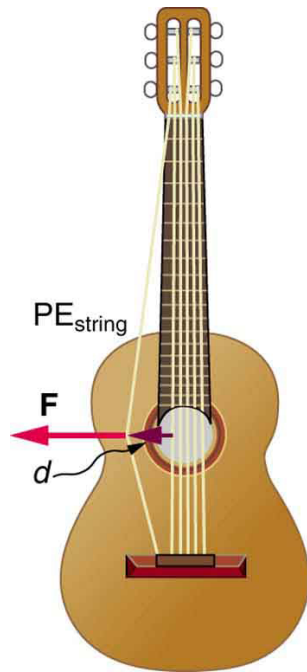
$$PE_s = \frac{1}{2}kx^2,$$

where k is the spring's force constant and x is the displacement from its undeformed position. The potential energy represents the work done *on* the spring and the energy stored in it as a result of stretching or compressing it a distance x . The potential energy of the spring PE_s does not depend on the path taken; it depends only on the stretch or squeeze x in the final configuration.



- (a) An undeformed spring has no PE_s stored in it. (b) The force needed to stretch (or compress) the spring a distance x has a magnitude $F = kx$, and the work done to stretch (or compress) it is $\frac{1}{2} kx^2$. Because the force is conservative, this work is stored as potential energy (PE_s) in the spring, and it can be fully recovered. (c) A graph of F vs. x has a slope of k , and the area under the graph is $\frac{1}{2} kx^2$. Thus the work done or potential energy stored is $\frac{1}{2} kx^2$.

The equation $PE_s = \frac{1}{2} kx^2$ has general validity beyond the special case for which it was derived. Potential energy can be stored in any elastic medium by deforming it. Indeed, the general definition of **potential energy** is energy due to position, shape, or configuration. For shape or position deformations, stored energy is $PE_s = \frac{1}{2} kx^2$, where k is the force constant of the particular system and x is its deformation. Another example is seen in [\[link\]](#) for a guitar string.



Work is done
to deform the
guitar string,
giving it
potential
energy.

When
released, the
potential
energy is
converted to
kinetic
energy and
back to
potential as
the string
oscillates
back and
forth. A very
small
fraction is
dissipated as

sound
energy,
slowly
removing
energy from
the string.

Conservation of Mechanical Energy

Let us now consider what form the work-energy theorem takes when only conservative forces are involved. This will lead us to the conservation of energy principle. The work-energy theorem states that the net work done by all forces acting on a system equals its change in kinetic energy. In equation form, this is

Equation:

$$W_{\text{net}} = \frac{1}{2}mv^2 - \frac{1}{2}mv_0^2 = \Delta\text{KE}.$$

If only conservative forces act, then

Equation:

$$W_{\text{net}} = W_{\text{c}},$$

where W_{c} is the total work done by all conservative forces. Thus,

Equation:

$$W_{\text{c}} = \Delta\text{KE}.$$

Now, if the conservative force, such as the gravitational force or a spring force, does work, the system loses potential energy. That is, $W_{\text{c}} = -\Delta\text{PE}$. Therefore,

Equation:

$$-\Delta PE = \Delta KE$$

or

Equation:

$$\Delta KE + \Delta PE = 0.$$

This equation means that the total kinetic and potential energy is constant for any process involving only conservative forces. That is,

Equation:

$$KE + PE = \text{constant}$$

or

(conservative forces only),

$$KE_i + PE_i = KE_f + PE_f$$

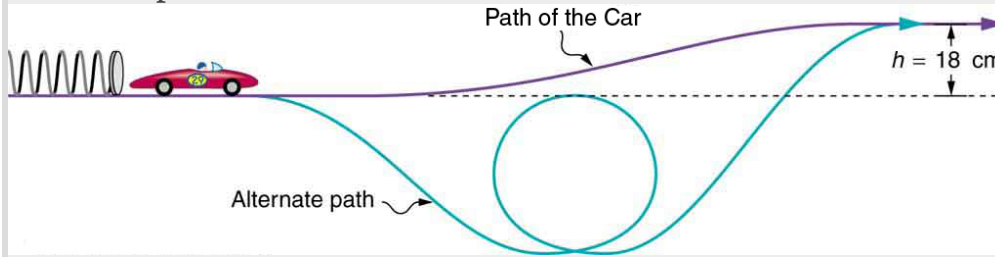
where i and f denote initial and final values. This equation is a form of the work-energy theorem for conservative forces; it is known as the **conservation of mechanical energy** principle. Remember that this applies to the extent that all the forces are conservative, so that friction is negligible. The total kinetic plus potential energy of a system is defined to be its **mechanical energy**, $(KE + PE)$. In a system that experiences only conservative forces, there is a potential energy associated with each force, and the energy only changes form between KE and the various types of PE, with the total energy remaining constant.

Example:

Using Conservation of Mechanical Energy to Calculate the Speed of a Toy Car

A 0.100-kg toy car is propelled by a compressed spring, as shown in [\[link\]](#). The car follows a track that rises 0.180 m above the starting point. The spring is compressed 4.00 cm and has a force constant of 250.0 N/m. Assuming work done by friction to be negligible, find (a) how fast the car

is going before it starts up the slope and (b) how fast it is going at the top of the slope.



A toy car is pushed by a compressed spring and coasts up a slope. Assuming negligible friction, the potential energy in the spring is first completely converted to kinetic energy, and then to a combination of kinetic and gravitational potential energy as the car rises. The details of the path are unimportant because all forces are conservative—the car would have the same final speed if it took the alternate path shown.

Strategy

The spring force and the gravitational force are conservative forces, so conservation of mechanical energy can be used. Thus,

Equation:

$$KE_i + PE_i = KE_f + PE_f$$

or

Equation:

$$\frac{1}{2}mv_i^2 + mgh_i + \frac{1}{2}kx_i^2 = \frac{1}{2}mv_f^2 + mgh_f + \frac{1}{2}kx_f^2,$$

where h is the height (vertical position) and x is the compression of the spring. This general statement looks complex but becomes much simpler when we start considering specific situations. First, we must identify the initial and final conditions in a problem; then, we enter them into the last equation to solve for an unknown.

Solution for (a)

This part of the problem is limited to conditions just before the car is released and just after it leaves the spring. Take the initial height to be zero, so that both h_i and h_f are zero. Furthermore, the initial speed v_i is zero and the final compression of the spring x_f is zero, and so several terms in the conservation of mechanical energy equation are zero and it simplifies to

Equation:

$$\frac{1}{2} k x_i^2 = \frac{1}{2} m v_f^2.$$

In other words, the initial potential energy in the spring is converted completely to kinetic energy in the absence of friction. Solving for the final speed and entering known values yields

Equation:

$$\begin{aligned} v_f &= \sqrt{\frac{k}{m}} x_i \\ &= \sqrt{\frac{250.0 \text{ N/m}}{0.100 \text{ kg}}} (0.0400 \text{ m}) \\ &= 2.00 \text{ m/s.} \end{aligned}$$

Solution for (b)

One method of finding the speed at the top of the slope is to consider conditions just before the car is released and just after it reaches the top of the slope, completely ignoring everything in between. Doing the same type of analysis to find which terms are zero, the conservation of mechanical energy becomes

Equation:

$$\frac{1}{2} k x_i^2 = \frac{1}{2} m v_f^2 + m g h_f.$$

This form of the equation means that the spring's initial potential energy is converted partly to gravitational potential energy and partly to kinetic energy. The final speed at the top of the slope will be less than at the bottom. Solving for v_f and substituting known values gives

Equation:

$$\begin{aligned}
 v_f &= \sqrt{\frac{kx_i^2}{m} - 2gh_f} \\
 &= \sqrt{\left(\frac{250.0 \text{ N/m}}{0.100 \text{ kg}}\right)(0.0400 \text{ m})^2 - 2(9.80 \text{ m/s}^2)(0.180 \text{ m})} \\
 &= 0.687 \text{ m/s.}
 \end{aligned}$$

Discussion

Another way to solve this problem is to realize that the car's kinetic energy before it goes up the slope is converted partly to potential energy—that is, to take the final conditions in part (a) to be the initial conditions in part (b).

Note that, for conservative forces, we do not directly calculate the work they do; rather, we consider their effects through their corresponding potential energies, just as we did in [\[link\]](#). Note also that we do not consider details of the path taken—only the starting and ending points are important (as long as the path is not impossible). This assumption is usually a tremendous simplification, because the path may be complicated and forces may vary along the way.

Note:

PhET Explorations: Energy Skate Park

Learn about conservation of energy with a skater dude! Build tracks, ramps and jumps for the skater and view the kinetic energy, potential energy and friction as he moves. You can also take the skater to different planets or even space!

https://phet.colorado.edu/sims/html/energy-skate-park-basics/latest/energy-skate-park-basics_en.html

Section Summary

- A conservative force is one for which work depends only on the starting and ending points of a motion, not on the path taken.
- We can define potential energy (PE) for any conservative force, just as we defined PE_g for the gravitational force.
- The potential energy of a spring is $PE_s = \frac{1}{2}kx^2$, where k is the spring's force constant and x is the displacement from its undeformed position.
- Mechanical energy is defined to be $KE + PE$ for a conservative force.
- When only conservative forces act on and within a system, the total mechanical energy is constant. In equation form,

Equation:

$$KE + PE = \text{constant}$$

or

$$KE_i + PE_i = KE_f + PE_f$$

where i and f denote initial and final values. This is known as the conservation of mechanical energy.

Conceptual Questions

Exercise:

Problem: What is a conservative force?

Exercise:

Problem:

The force exerted by a diving board is conservative, provided the internal friction is negligible. Assuming friction is negligible, describe changes in the potential energy of a diving board as a swimmer dives from it, starting just before the swimmer steps on the board until just after his feet leave it.

Exercise:

Problem:

Define mechanical energy. What is the relationship of mechanical energy to nonconservative forces? What happens to mechanical energy if only conservative forces act?

Exercise:**Problem:**

What is the relationship of potential energy to conservative force?

Problems & Exercises**Exercise:****Problem:**

A 5.00×10^5 -kg subway train is brought to a stop from a speed of 0.500 m/s in 0.400 m by a large spring bumper at the end of its track. What is the force constant k of the spring?

Solution:**Equation:**

$$7.81 \times 10^5 \text{ N/m}$$

Exercise:**Problem:**

A pogo stick has a spring with a force constant of 2.50×10^4 N/m, which can be compressed 12.0 cm. To what maximum height can a child jump on the stick using only the energy in the spring, if the child and stick have a total mass of 40.0 kg? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Energy](#).

Glossary

conservative force

a force that does the same work for any given initial and final configuration, regardless of the path followed

potential energy

energy due to position, shape, or configuration

potential energy of a spring

the stored energy of a spring as a function of its displacement; when Hooke's law applies, it is given by the expression $\frac{1}{2}kx^2$ where x is the distance the spring is compressed or extended and k is the spring constant

conservation of mechanical energy

the rule that the sum of the kinetic energies and potential energies remains constant if only conservative forces act on and within a system

mechanical energy

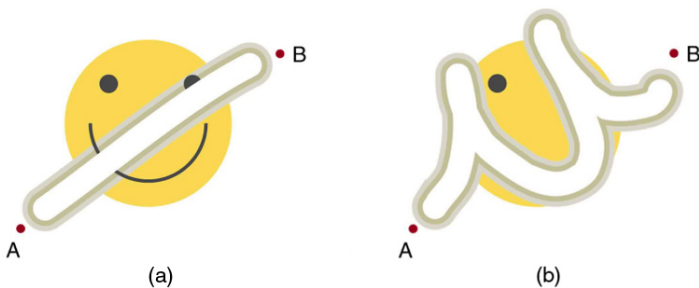
the sum of kinetic energy and potential energy

Nonconservative Forces

- Define nonconservative forces and explain how they affect mechanical energy.
- Show how the principle of conservation of energy can be applied by treating the conservative forces in terms of their potential energies and any nonconservative forces in terms of the work they do.

Nonconservative Forces and Friction

Forces are either conservative or nonconservative. Conservative forces were discussed in [Conservative Forces and Potential Energy](#). A **nonconservative force** is one for which work depends on the path taken. Friction is a good example of a nonconservative force. As illustrated in [\[link\]](#), work done against friction depends on the length of the path between the starting and ending points. Because of this dependence on path, there is no potential energy associated with nonconservative forces. An important characteristic is that the work done by a nonconservative force *adds or removes mechanical energy from a system*. **Friction**, for example, creates **thermal energy** that dissipates, removing energy from the system. Furthermore, even if the thermal energy is retained or captured, it cannot be fully converted back to work, so it is lost or not recoverable in that sense as well.

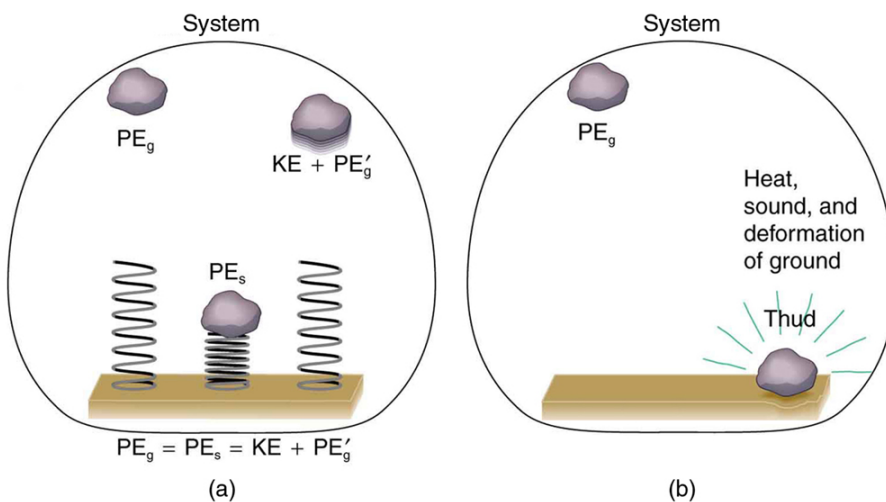


The amount of the happy face erased depends on the path taken by the eraser between points A and B, as does the work done against friction. Less work is done and less of the face

is erased for the path in (a) than for the path in (b). The force here is friction, and most of the work goes into thermal energy that subsequently leaves the system (the happy face plus the eraser). The energy expended cannot be fully recovered.

How Nonconservative Forces Affect Mechanical Energy

Mechanical energy may not be conserved when nonconservative forces act. For example, when a car is brought to a stop by friction on level ground, it loses kinetic energy, which is dissipated as thermal energy, reducing its mechanical energy. [\[link\]](#) compares the effects of conservative and nonconservative forces. We often choose to understand simpler systems such as that described in [\[link\]](#)(a) first before studying more complicated systems as in [\[link\]](#)(b).



Comparison of the effects of conservative and nonconservative forces on the mechanical energy of a system. (a) A system with only conservative

forces. When a rock is dropped onto a spring, its mechanical energy remains constant (neglecting air resistance) because the force in the spring is conservative. The spring can propel the rock back to its original height, where it once again has only potential energy due to gravity. (b) A system with nonconservative forces. When the same rock is dropped onto the ground, it is stopped by nonconservative forces that dissipate its mechanical energy as thermal energy, sound, and surface distortion. The rock has lost mechanical energy.

How the Work-Energy Theorem Applies

Now let us consider what form the work-energy theorem takes when both conservative and nonconservative forces act. We will see that the work done by nonconservative forces equals the change in the mechanical energy of a system. As noted in [Kinetic Energy and the Work-Energy Theorem](#), the work-energy theorem states that the net work on a system equals the change in its kinetic energy, or $W_{\text{net}} = \Delta\text{KE}$. The net work is the sum of the work by nonconservative forces plus the work by conservative forces. That is,

Equation:

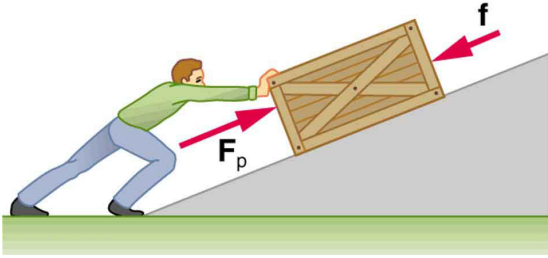
$$W_{\text{net}} = W_{\text{nc}} + W_{\text{c}},$$

so that

Equation:

$$W_{\text{nc}} + W_{\text{c}} = \Delta\text{KE},$$

where W_{nc} is the total work done by all nonconservative forces and W_{c} is the total work done by all conservative forces.



A person pushes a crate up a ramp, doing work on the crate. Friction and gravitational force (not shown) also do work on the crate; both forces oppose the person's push. As the crate is pushed up the ramp, it gains mechanical energy, implying that the work done by the person is greater than the work done by friction.

Consider [\[link\]](#), in which a person pushes a crate up a ramp and is opposed by friction. As in the previous section, we note that work done by a conservative force comes from a loss of gravitational potential energy, so that $W_c = -\Delta PE$. Substituting this equation into the previous one and solving for W_{nc} gives

Equation:

$$W_{nc} = \Delta KE + \Delta PE.$$

This equation means that the total mechanical energy ($KE + PE$) changes by exactly the amount of work done by nonconservative forces. In [\[link\]](#), this is the work done by the person minus the work done by friction. So even if energy is not conserved for the system of interest (such as the crate), we know that an equal amount of work was done to cause the change in total mechanical energy.

We rearrange $W_{\text{nc}} = \Delta\text{KE} + \Delta\text{PE}$ to obtain

Equation:

$$\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f.$$

This means that the amount of work done by nonconservative forces adds to the mechanical energy of a system. If W_{nc} is positive, then mechanical energy is increased, such as when the person pushes the crate up the ramp in [\[link\]](#). If W_{nc} is negative, then mechanical energy is decreased, such as when the rock hits the ground in [\[link\]](#)(b). If W_{nc} is zero, then mechanical energy is conserved, and nonconservative forces are balanced. For example, when you push a lawn mower at constant speed on level ground, your work done is removed by the work of friction, and the mower has a constant energy.

Applying Energy Conservation with Nonconservative Forces

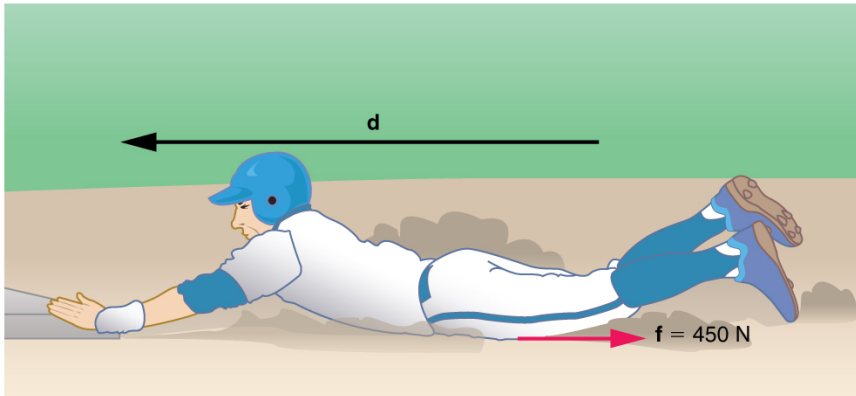
When no change in potential energy occurs, applying $\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f$ amounts to applying the work-energy theorem by setting the change in kinetic energy to be equal to the net work done on the system, which in the most general case includes both conservative and nonconservative forces. But when seeking instead to find a change in total mechanical energy in situations that involve changes in both potential and kinetic energy, the previous equation $\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f$ says that you can start by finding the change in mechanical energy that would have resulted from just the conservative forces, including the potential energy changes, and add to it the work done, with the proper sign, by any nonconservative forces involved.

Example:

Calculating Distance Traveled: How Far a Baseball Player Slides

Consider the situation shown in [\[link\]](#), where a baseball player slides to a stop on level ground. Using energy considerations, calculate the distance

the 65.0-kg baseball player slides, given that his initial speed is 6.00 m/s and the force of friction against him is a constant 450 N.



The baseball player slides to a stop in a distance d . In the process, friction removes the player's kinetic energy by doing an amount of work fd equal to the initial kinetic energy.

Strategy

Friction stops the player by converting his kinetic energy into other forms, including thermal energy. In terms of the work-energy theorem, the work done by friction, which is negative, is added to the initial kinetic energy to reduce it to zero. The work done by friction is negative, because $\mathbf{f}$ is in the opposite direction of the motion (that is, $\theta = 180^\circ$, and so $\cos \theta = -1$). Thus $W_{\text{nc}} = -fd$. The equation simplifies to

Equation:

$$\frac{1}{2}mv_i^2 - fd = 0$$

or

Equation:

$$fd = \frac{1}{2}mv_i^2.$$

This equation can now be solved for the distance d .

Solution

Solving the previous equation for d and substituting known values yields
Equation:

$$\begin{aligned}d &= \frac{mv_i^2}{2f} \\&= \frac{(65.0 \text{ kg})(6.00 \text{ m/s})^2}{(2)(450 \text{ N})} \\&= 2.60 \text{ m.}\end{aligned}$$

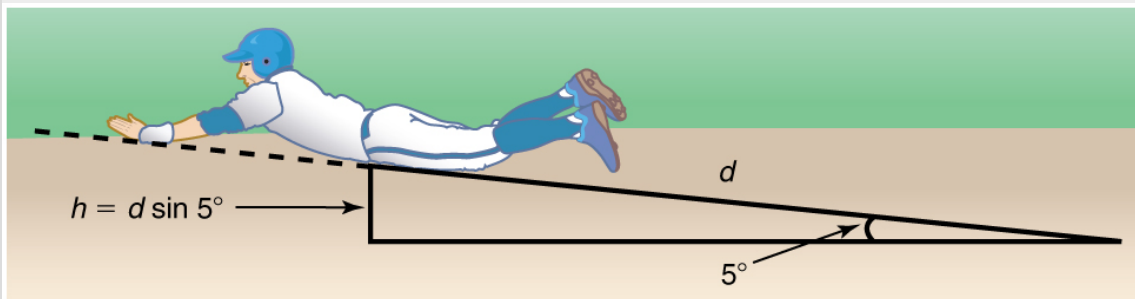
Discussion

The most important point of this example is that the amount of nonconservative work equals the change in mechanical energy. For example, you must work harder to stop a truck, with its large mechanical energy, than to stop a mosquito.

Example:

Calculating Distance Traveled: Sliding Up an Incline

Suppose that the player from [\[link\]](#) is running up a hill having a 5.00° incline upward with a surface similar to that in the baseball stadium. The player slides with the same initial speed, and the frictional force is still 450 N. Determine how far he slides.



The same baseball player slides to a stop on a 5.00° slope.

Strategy

In this case, the work done by the nonconservative friction force on the player reduces the mechanical energy he has from his kinetic energy at zero height, to the final mechanical energy he has by moving through

distance d to reach height h along the hill, with $h = d \sin 5.00^\circ$. This is expressed by the equation

Equation:

$$\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f.$$

Solution

The work done by friction is again $W_{\text{nc}} = -fd$; initially the potential energy is $\text{PE}_i = mg \cdot 0 = 0$ and the kinetic energy is $\text{KE}_i = \frac{1}{2}mv_i^2$; the final energy contributions are $\text{KE}_f = 0$ for the kinetic energy and $\text{PE}_f = mgh = mgd \sin \theta$ for the potential energy.

Substituting these values gives

Equation:

$$\frac{1}{2}mv_i^2 + 0 + (-fd) = 0 + mgd \sin \theta.$$

Solve this for d to obtain

Equation:

$$\begin{aligned} d &= \frac{(\frac{1}{2})mv_i^2}{f + mg \sin \theta} \\ &= \frac{(0.5)(65.0 \text{ kg})(6.00 \text{ m/s})^2}{450 \text{ N} + (65.0 \text{ kg})(9.80 \text{ m/s}^2) \sin (5.00^\circ)} \\ &= 2.31 \text{ m.} \end{aligned}$$

Discussion

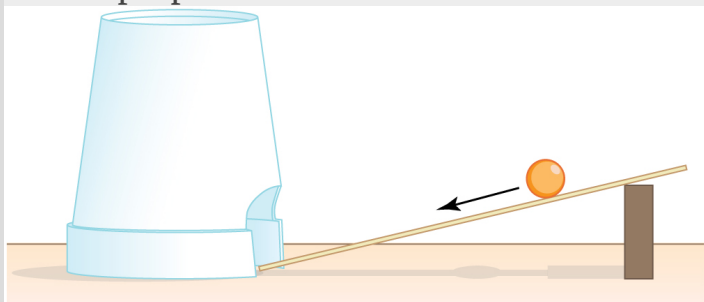
As might have been expected, the player slides a shorter distance by sliding uphill. Note that the problem could also have been solved in terms of the forces directly and the work energy theorem, instead of using the potential energy. This method would have required combining the normal force and force of gravity vectors, which no longer cancel each other because they point in different directions, and friction, to find the net force. You could then use the net force and the net work to find the distance d that reduces the kinetic energy to zero. By applying conservation of energy and using the potential energy instead, we need only consider the gravitational potential energy mgh , without combining and resolving force vectors. This simplifies the solution considerably.

Note:**Making Connections: Take-Home Investigation—Determining Friction from the Stopping Distance**

This experiment involves the conversion of gravitational potential energy into thermal energy. Use the ruler, book, and marble from [Take-Home Investigation—Converting Potential to Kinetic Energy](#). In addition, you will need a foam cup with a small hole in the side, as shown in [\[link\]](#). From the 10-cm position on the ruler, let the marble roll into the cup positioned at the bottom of the ruler. Measure the distance d the cup moves before stopping. What forces caused it to stop? What happened to the kinetic energy of the marble at the bottom of the ruler? Next, place the marble at the 20-cm and the 30-cm positions and again measure the distance the cup moves after the marble enters it. Plot the distance the cup moves versus the initial marble position on the ruler. Is this relationship linear?

With some simple assumptions, you can use these data to find the coefficient of kinetic friction μ_k of the cup on the table. The force of friction f on the cup is $\mu_k N$, where the normal force N is just the weight of the cup plus the marble. The normal force and force of gravity do no work because they are perpendicular to the displacement of the cup, which moves horizontally. The work done by friction is fd . You will need the mass of the marble as well to calculate its initial kinetic energy.

It is interesting to do the above experiment also with a steel marble (or ball bearing). Releasing it from the same positions on the ruler as you did with the glass marble, is the velocity of this steel marble the same as the velocity of the marble at the bottom of the ruler? Is the distance the cup moves proportional to the mass of the steel and glass marbles?



Rolling a marble down a ruler into a foam cup.

Note:**PhET Explorations: The Ramp**

Explore forces, energy and work as you push household objects up and down a ramp. Lower and raise the ramp to see how the angle of inclination affects the parallel forces acting on the file cabinet. Graphs show forces, energy and work. [Click to open media in new browser.](#)

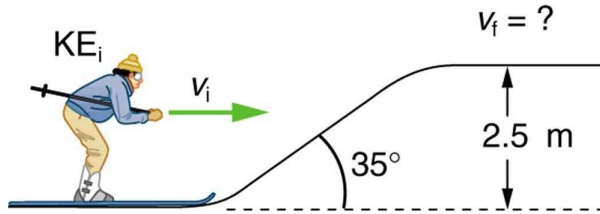
Section Summary

- A nonconservative force is one for which work depends on the path.
- Friction is an example of a nonconservative force that changes mechanical energy into thermal energy.
- Work W_{nc} done by a nonconservative force changes the mechanical energy of a system. In equation form, $W_{\text{nc}} = \Delta\text{KE} + \Delta\text{PE}$ or, equivalently, $\text{KE}_i + \text{PE}_i + W_{\text{nc}} = \text{KE}_f + \text{PE}_f$.
- When both conservative and nonconservative forces act, energy conservation can be applied and used to calculate motion in terms of the known potential energies of the conservative forces and the work done by nonconservative forces, instead of finding the net work from the net force, or having to directly apply Newton's laws.

Problems & Exercises

Exercise:**Problem:**

A 60.0-kg skier with an initial speed of 12.0 m/s coasts up a 2.50-m-high rise as shown in [\[link\]](#). Find her final speed at the top, given that the coefficient of friction between her skis and the snow is 0.0800. (Hint: Find the distance traveled up the incline assuming a straight-line path as shown in the figure.)



The skier's initial kinetic energy is partially used in coasting to the top of a rise.

Solution:

9.46 m/s

Exercise:

Problem:

(a) How high a hill can a car coast up (engine disengaged) if work done by friction is negligible and its initial speed is 110 km/h? (b) If, in actuality, a 750-kg car with an initial speed of 110 km/h is observed to coast up a hill to a height 22.0 m above its starting point, how much thermal energy was generated by friction? (c) What is the average force of friction if the hill has a slope 2.5° above the horizontal?

Glossary

nonconservative force

a force whose work depends on the path followed between the given initial and final configurations

friction

the force between surfaces that opposes one sliding on the other; friction changes mechanical energy into thermal energy

Conservation of Energy

- Explain the law of the conservation of energy.
- Describe some of the many forms of energy.
- Define efficiency of an energy conversion process as the fraction left as useful energy or work, rather than being transformed, for example, into thermal energy.

Law of Conservation of Energy

Energy, as we have noted, is conserved, making it one of the most important physical quantities in nature. The **law of conservation of energy** can be stated as follows:

Total energy is constant in any process. It may change in form or be transferred from one system to another, but the total remains the same.

We have explored some forms of energy and some ways it can be transferred from one system to another. This exploration led to the definition of two major types of energy—mechanical energy ($KE + PE$) and energy transferred via work done by nonconservative forces (W_{nc}). But energy takes *many* other forms, manifesting itself in *many* different ways, and we need to be able to deal with all of these before we can write an equation for the above general statement of the conservation of energy.

Other Forms of Energy than Mechanical Energy

At this point, we deal with all other forms of energy by lumping them into a single group called other energy (OE). Then we can state the conservation of energy in equation form as

Equation:

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f.$$

All types of energy and work can be included in this very general statement of conservation of energy. Kinetic energy is KE , work done by a conservative force is represented by PE , work done by nonconservative forces is W_{nc} , and

all other energies are included as OE. This equation applies to all previous examples; in those situations OE was constant, and so it subtracted out and was not directly considered.

Note:

Making Connections: Usefulness of the Energy Conservation Principle

The fact that energy is conserved and has many forms makes it very important. You will find that energy is discussed in many contexts, because it is involved in all processes. It will also become apparent that many situations are best understood in terms of energy and that problems are often most easily conceptualized and solved by considering energy.

When does OE play a role? One example occurs when a person eats. Food is oxidized with the release of carbon dioxide, water, and energy. Some of this chemical energy is converted to kinetic energy when the person moves, to potential energy when the person changes altitude, and to thermal energy (another form of OE).

Some of the Many Forms of Energy

What are some other forms of energy? You can probably name a number of forms of energy not yet discussed. Many of these will be covered in later chapters, but let us detail a few here. **Electrical energy** is a common form that is converted to many other forms and does work in a wide range of practical situations. Fuels, such as gasoline and food, carry **chemical energy** that can be transferred to a system through oxidation. Chemical fuel can also produce electrical energy, such as in batteries. Batteries can in turn produce light, which is a very pure form of energy. Most energy sources on Earth are in fact stored energy from the energy we receive from the Sun. We sometimes refer to this as **radiant energy**, or electromagnetic radiation, which includes visible light, infrared, and ultraviolet radiation. **Nuclear energy** comes from processes that convert measurable amounts of mass into energy. Nuclear energy is transformed into the energy of sunlight, into electrical energy in power plants, and into the energy of the heat transfer and blast in weapons.

Atoms and molecules inside all objects are in random motion. This internal mechanical energy from the random motions is called **thermal energy**, because it is related to the temperature of the object. These and all other forms of energy can be converted into one another and can do work.

[\[link\]](#) gives the amount of energy stored, used, or released from various objects and in various phenomena. The range of energies and the variety of types and situations is impressive.

Note:

Problem-Solving Strategies for Energy

You will find the following problem-solving strategies useful whenever you deal with energy. The strategies help in organizing and reinforcing energy concepts. In fact, they are used in the examples presented in this chapter. The familiar general problem-solving strategies presented earlier—involving identifying physical principles, knowns, and unknowns, checking units, and so on—continue to be relevant here.

Step 1. Determine the system of interest and identify what information is given and what quantity is to be calculated. A sketch will help.

Step 2. Examine all the forces involved and determine whether you know or are given the potential energy from the work done by the forces. Then use step 3 or step 4.

Step 3. If you know the potential energies for the forces that enter into the problem, then forces are all conservative, and you can apply conservation of mechanical energy simply in terms of potential and kinetic energy. The equation expressing conservation of energy is

Equation:

$$KE_i + PE_i = KE_f + PE_f.$$

Step 4. If you know the potential energy for only some of the forces, possibly because some of them are nonconservative and do not have a potential energy, or if there are other energies that are not easily treated in terms of force and work, then the conservation of energy law in its most general form must be used.

Equation:

$$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f.$$

In most problems, one or more of the terms is zero, simplifying its solution. Do not calculate W_c , the work done by conservative forces; it is already incorporated in the PE terms.

Step 5. You have already identified the types of work and energy involved (in step 2). Before solving for the unknown, *eliminate terms wherever possible* to simplify the algebra. For example, choose $h = 0$ at either the initial or final point, so that PE_g is zero there. Then solve for the unknown in the customary manner.

Step 6. *Check the answer to see if it is reasonable.* Once you have solved a problem, reexamine the forms of work and energy to see if you have set up the conservation of energy equation correctly. For example, work done against friction should be negative, potential energy at the bottom of a hill should be less than that at the top, and so on. Also check to see that the numerical value obtained is reasonable. For example, the final speed of a skateboarder who coasts down a 3-m-high ramp could reasonably be 20 km/h, but *not* 80 km/h.

Transformation of Energy

The transformation of energy from one form into others is happening all the time. The chemical energy in food is converted into thermal energy through metabolism; light energy is converted into chemical energy through photosynthesis. In a larger example, the chemical energy contained in coal is converted into thermal energy as it burns to turn water into steam in a boiler. This thermal energy in the steam in turn is converted to mechanical energy as it spins a turbine, which is connected to a generator to produce electrical energy. (In all of these examples, not all of the initial energy is converted into the forms mentioned. This important point is discussed later in this section.)

Another example of energy conversion occurs in a solar cell. Sunlight impinging on a solar cell (see [\[link\]](#)) produces electricity, which in turn can be used to run an electric motor. Energy is converted from the primary source of solar energy into electrical energy and then into mechanical energy.



Solar energy is converted into electrical energy by solar cells, which is used to run a motor in this solar-power aircraft. (credit: NASA)

Object/phenomenon	Energy in joules
Big Bang	10^{68}
Energy released in a supernova	10^{44}
Fusion of all the hydrogen in Earth's oceans	10^{34}
Annual world energy use	4×10^{20}

Object/phenomenon	Energy in joules
Large fusion bomb (9 megaton)	3.8×10^{16}
1 kg hydrogen (fusion to helium)	6.4×10^{14}
1 kg uranium (nuclear fission)	8.0×10^{13}
Hiroshima-size fission bomb (10 kiloton)	4.2×10^{13}
90,000-metric ton aircraft carrier at 30 knots	1.1×10^{10}
1 barrel crude oil	5.9×10^9
1 ton TNT	4.2×10^9
1 gallon of gasoline	1.2×10^8
Daily home electricity use (developed countries)	7×10^7
Daily adult food intake (recommended)	1.2×10^7

Object/phenomenon	Energy in joules
1000-kg car at 90 km/h	3.1×10^5
1 g fat (9.3 kcal)	3.9×10^4
ATP hydrolysis reaction	3.2×10^4
1 g carbohydrate (4.1 kcal)	1.7×10^4
1 g protein (4.1 kcal)	1.7×10^4
Tennis ball at 100 km/h	22
Mosquito (10^{-2} g at 0.5 m/s)	1.3×10^{-6}
Single electron in a TV tube beam	4.0×10^{-15}
Energy to break one DNA strand	10^{-19}

Energy of Various Objects and Phenomena

Efficiency

Even though energy is conserved in an energy conversion process, the output of *useful energy* or work will be less than the energy input. The **efficiency** Eff of an energy conversion process is defined as

Equation:

$$\text{Efficiency}(\text{Eff}) = \frac{\text{useful energy or work output}}{\text{total energy input}} = \frac{W_{\text{out}}}{E_{\text{in}}}.$$

[\[link\]](#) lists some efficiencies of mechanical devices and human activities. In a coal-fired power plant, for example, about 40% of the chemical energy in the coal becomes useful electrical energy. The other 60% transforms into other (perhaps less useful) energy forms, such as thermal energy, which is then released to the environment through combustion gases and cooling towers.

Activity/device	Efficiency (%) [footnote] Representative values
Cycling and climbing	20
Swimming, surface	2
Swimming, submerged	4
Shoveling	3
Weightlifting	9
Steam engine	17
Gasoline engine	30

Activity/device	Efficiency (%) ^[footnote] Representative values
Diesel engine	35
Nuclear power plant	35
Coal power plant	42
Electric motor	98
Compact fluorescent light	20
Gas heater (residential)	90
Solar cell	10

Efficiency of the Human Body and Mechanical Devices

Note:

PhET Explorations: Masses and Springs

A realistic mass and spring laboratory. Hang masses from springs and adjust the spring stiffness and damping. You can even slow time. Transport the lab to different planets. A chart shows the kinetic, potential, and thermal energies for each spring.

https://phet.colorado.edu/sims/mass-spring-lab/mass-spring-lab_en.html

Section Summary

- The law of conservation of energy states that the total energy is constant in any process. Energy may change in form or be transferred from one system to another, but the total remains the same.
- When all forms of energy are considered, conservation of energy is written in equation form as

$KE_i + PE_i + W_{nc} + OE_i = KE_f + PE_f + OE_f$, where OE is all **other forms of energy** besides mechanical energy.

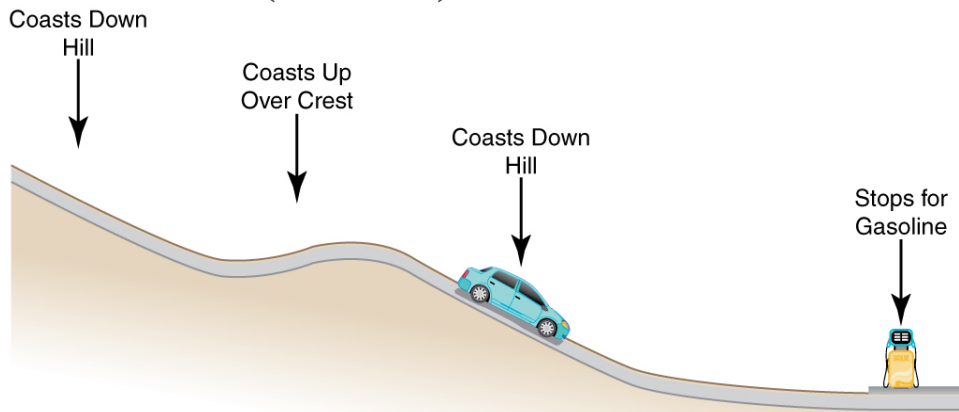
- Commonly encountered forms of energy include electric energy, chemical energy, radiant energy, nuclear energy, and thermal energy.
- Energy is often utilized to do work, but it is not possible to convert all the energy of a system to work.
- The efficiency Eff of a machine or human is defined to be $Eff = \frac{W_{out}}{E_{in}}$, where W_{out} is useful work output and E_{in} is the energy consumed.

Conceptual Questions

Exercise:

Problem:

Consider the following scenario. A car for which friction is *not* negligible accelerates from rest down a hill, running out of gasoline after a short distance. The driver lets the car coast farther down the hill, then up and over a small crest. He then coasts down that hill into a gas station, where he brakes to a stop and fills the tank with gasoline. Identify the forms of energy the car has, and how they are changed and transferred in this series of events. (See [\[link\]](#).)



A car experiencing non-negligible friction coasts down a hill, over a small crest, then downhill again, and comes to a stop at a gas station.

Exercise:**Problem:**

Describe the energy transfers and transformations for a javelin, starting from the point at which an athlete picks up the javelin and ending when the javelin is stuck into the ground after being thrown.

Exercise:**Problem:**

Do devices with efficiencies of less than one violate the law of conservation of energy? Explain.

Exercise:**Problem:**

List four different forms or types of energy. Give one example of a conversion from each of these forms to another form.

Exercise:

Problem: List the energy conversions that occur when riding a bicycle.

Problems & Exercises**Exercise:****Problem:**

Using values from [\[link\]](#), how many DNA molecules could be broken by the energy carried by a single electron in the beam of an old-fashioned TV tube? (These electrons were not dangerous in themselves, but they did create dangerous x rays. Later model tube TVs had shielding that absorbed x rays before they escaped and exposed viewers.)

Solution:

4×10^4 molecules

Exercise:**Problem:**

Using energy considerations and assuming negligible air resistance, show that a rock thrown from a bridge 20.0 m above water with an initial speed of 15.0 m/s strikes the water with a speed of 24.8 m/s independent of the direction thrown.

Solution:

Equating ΔPE_g and ΔKE , we obtain

$$v = \sqrt{2gh + v_0^2} = \sqrt{2(9.80 \text{ m/s}^2)(20.0 \text{ m}) + (15.0 \text{ m/s})^2} = 24.8 \text{ m/s}$$

Exercise:**Problem:**

If the energy in fusion bombs were used to supply the energy needs of the world, how many of the 9-megaton variety would be needed for a year's supply of energy (using data from [\[link\]](#))? This is not as far-fetched as it may sound—there are thousands of nuclear bombs, and their energy can be trapped in underground explosions and converted to electricity, as natural geothermal energy is.

Exercise:**Problem:**

(a) Use of hydrogen fusion to supply energy is a dream that may be realized in the next century. Fusion would be a relatively clean and almost limitless supply of energy, as can be seen from [\[link\]](#). To illustrate this, calculate how many years the present energy needs of the world could be supplied by one millionth of the oceans' hydrogen fusion energy. (b) How does this time compare with historically significant events, such as the duration of stable economic systems?

Solution:

(a) 25×10^6 years

(b) This is much, much longer than human time scales.

Glossary

law of conservation of energy

the general law that total energy is constant in any process; energy may change in form or be transferred from one system to another, but the total remains the same

electrical energy

the energy carried by a flow of charge

chemical energy

the energy in a substance stored in the bonds between atoms and molecules that can be released in a chemical reaction

radiant energy

the energy carried by electromagnetic waves

nuclear energy

energy released by changes within atomic nuclei, such as the fusion of two light nuclei or the fission of a heavy nucleus

thermal energy

the energy within an object due to the random motion of its atoms and molecules that accounts for the object's temperature

efficiency

a measure of the effectiveness of the input of energy to do work; useful energy or work divided by the total input of energy

Power

- Calculate power by calculating changes in energy over time.
- Examine power consumption and calculations of the cost of energy consumed.

What is Power?

Power—the word conjures up many images: a professional football player muscling aside his opponent, a dragster roaring away from the starting line, a volcano blowing its lava into the atmosphere, or a rocket blasting off, as in [\[link\]](#).



This powerful rocket on the Space Shuttle *Endeavor* did work and consumed energy at a very high rate. (credit: NASA)

These images of power have in common the rapid performance of work, consistent with the scientific definition of **power** (P) as the rate at which work is done.

Note:**Power**

Power is the rate at which work is done.

Equation:

$$P = \frac{W}{t}$$

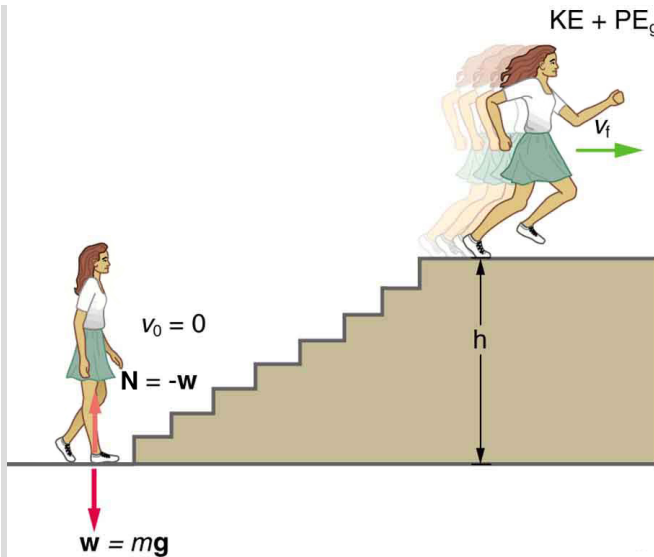
The SI unit for power is the **watt** (W), where 1 watt equals 1 joule/second ($1 \text{ W} = 1 \text{ J/s}$).

Because work is energy transfer, power is also the rate at which energy is expended. A 60-W light bulb, for example, expends 60 J of energy per second. Great power means a large amount of work or energy developed in a short time. For example, when a powerful car accelerates rapidly, it does a large amount of work and consumes a large amount of fuel in a short time.

Calculating Power from Energy

Example:**Calculating the Power to Climb Stairs**

What is the power output for a 60.0-kg woman who runs up a 3.00 m high flight of stairs in 3.50 s, starting from rest but having a final speed of 2.00 m/s? (See [\[link\]](#).)



When this woman runs upstairs starting from rest, she converts the chemical energy originally from food into kinetic energy and gravitational potential energy. Her power output depends on how fast she does this.

Strategy and Concept

The work going into mechanical energy is $W = KE + PE$. At the bottom of the stairs, we take both KE and PE_g as initially zero; thus,

$W = KE_f + PE_g = \frac{1}{2}mv_f^2 + mgh$, where h is the vertical height of the stairs. Because all terms are given, we can calculate W and then divide it by time to get power.

Solution

Substituting the expression for W into the definition of power given in the previous equation, $P = W/t$ yields

Equation:

$$P = \frac{W}{t} = \frac{\frac{1}{2}mv_f^2 + mgh}{t}.$$

Entering known values yields

Equation:

$$\begin{aligned} P &= \frac{0.5(60.0 \text{ kg})(2.00 \text{ m/s})^2 + (60.0 \text{ kg})(9.80 \text{ m/s}^2)(3.00 \text{ m})}{3.50 \text{ s}} \\ &= \frac{120 \text{ J} + 1764 \text{ J}}{3.50 \text{ s}} \\ &= 538 \text{ W}. \end{aligned}$$

Discussion

The woman does 1764 J of work to move up the stairs compared with only 120 J to increase her kinetic energy; thus, most of her power output is required for climbing rather than accelerating.

It is impressive that this woman's useful power output is slightly less than 1 **horsepower** (1 hp = 746 W)! People can generate more than a horsepower with their leg muscles for short periods of time by rapidly converting available blood sugar and oxygen into work output. (A horse can put out 1 hp for hours on end.) Once oxygen is depleted, power output decreases and the person begins to breathe rapidly to obtain oxygen to metabolize more food—this is known as the *aerobic* stage of exercise. If the woman climbed the stairs slowly, then her power output would be much less, although the amount of work done would be the same.

Note:**Making Connections: Take-Home Investigation—Measure Your Power Rating**

Determine your own power rating by measuring the time it takes you to climb a flight of stairs. We will ignore the gain in kinetic energy, as the above example showed that it was a small portion of the energy gain. Don't expect that your output will be more than about 0.5 hp.

Examples of Power

Examples of power are limited only by the imagination, because there are as many types as there are forms of work and energy. (See [\[link\]](#) for some examples.) Sunlight reaching Earth's surface carries a maximum power of about 1.3 kilowatts per square meter (kW/m^2). A tiny fraction of this is retained by Earth over the long term. Our consumption rate of fossil fuels is far greater than the rate at which they are stored, so it is inevitable that they will be depleted. Power implies that energy is transferred, perhaps changing form. It is never possible to change one form completely into another without losing some of it as thermal energy. For example, a 60-W incandescent bulb converts only 5 W of electrical power to light, with 55 W dissipating into thermal energy. Furthermore, the typical electric power plant converts only 35 to 40% of its fuel into electricity. The remainder becomes a huge amount of thermal energy that must be dispersed as heat transfer, as rapidly as it is created. A coal-fired power plant may produce 1000 megawatts; 1 megawatt (MW) is 10^6 W of electric power. But the power plant consumes chemical energy at a rate of about 2500 MW, creating heat transfer to the surroundings at a rate of 1500 MW. (See [\[link\]](#).)



Tremendous amounts of electric power are generated by coal-fired power plants such as this one in China, but an even larger amount of power goes into heat transfer to the surroundings.

The large cooling towers here are needed to transfer heat as rapidly as it is produced. The transfer of heat is not unique to coal plants but is an unavoidable consequence of generating electric power from any fuel—nuclear, coal, oil, natural gas, or the like. (credit: Kleinolive, Wikimedia Commons)

Object or Phenomenon	Power in Watts
Supernova (at peak)	5×10^{37}
Milky Way galaxy	10^{37}
Crab Nebula pulsar	10^{28}
The Sun	4×10^{26}

Object or Phenomenon	Power in Watts
Volcanic eruption (maximum)	4×10^{15}
Lightning bolt	2×10^{12}
Nuclear power plant (total electric and heat transfer)	3×10^9
Aircraft carrier (total useful and heat transfer)	10^8
Dragster (total useful and heat transfer)	2×10^6
Car (total useful and heat transfer)	8×10^4
Football player (total useful and heat transfer)	5×10^3
Clothes dryer	4×10^3
Person at rest (all heat transfer)	100

Object or Phenomenon	Power in Watts
Typical incandescent light bulb (total useful and heat transfer)	60
Heart, person at rest (total useful and heat transfer)	8
Electric clock	3
Pocket calculator	10^{-3}

Power Output or Consumption

Power and Energy Consumption

We usually have to pay for the energy we use. It is interesting and easy to estimate the cost of energy for an electrical appliance if its power consumption rate and time used are known. The higher the power consumption rate and the longer the appliance is used, the greater the cost of that appliance. The power consumption rate is $P = W/t = E/t$, where E is the energy supplied by the electricity company. So the energy consumed over a time t is

Equation:

$$E = Pt.$$

Electricity bills state the energy used in units of **kilowatt-hours** ($\text{kW} \cdot \text{h}$), which is the product of power in kilowatts and time in hours. This unit is convenient because electrical power consumption at the kilowatt level for hours at a time is typical.

Example:**Calculating Energy Costs**

What is the cost of running a 0.200-kW computer 6.00 h per day for 30.0 d if the cost of electricity is \$0.120 per kW · h?

Strategy

Cost is based on energy consumed; thus, we must find E from $E = Pt$ and then calculate the cost. Because electrical energy is expressed in kW · h, at the start of a problem such as this it is convenient to convert the units into kW and hours.

Solution

The energy consumed in kW · h is

Equation:

$$\begin{aligned} E &= Pt = (0.200 \text{ kW})(6.00 \text{ h/d})(30.0 \text{ d}) \\ &= 36.0 \text{ kW} \cdot \text{h}, \end{aligned}$$

and the cost is simply given by

Equation:

$$\text{cost} = (36.0 \text{ kW} \cdot \text{h})(\$0.120 \text{ per kW} \cdot \text{h}) = \$4.32 \text{ per month.}$$

Discussion

The cost of using the computer in this example is neither exorbitant nor negligible. It is clear that the cost is a combination of power and time. When both are high, such as for an air conditioner in the summer, the cost is high.

The motivation to save energy has become more compelling with its ever-increasing price. Armed with the knowledge that energy consumed is the product of power and time, you can estimate costs for yourself and make the necessary value judgments about where to save energy. Either power or time must be reduced. It is most cost-effective to limit the use of high-power devices that normally operate for long periods of time, such as water heaters and air conditioners. This would not include relatively high power devices like toasters, because they are on only a few minutes per day. It would also not include electric clocks, in spite of their 24-hour-per-day

usage, because they are very low power devices. It is sometimes possible to use devices that have greater efficiencies—that is, devices that consume less power to accomplish the same task. One example is the compact fluorescent light bulb, which produces over four times more light per watt of power consumed than its incandescent cousin.

Modern civilization depends on energy, but current levels of energy consumption and production are not sustainable. The likelihood of a link between global warming and fossil fuel use (with its concomitant production of carbon dioxide), has made reduction in energy use as well as a shift to non-fossil fuels of the utmost importance. Even though energy in an isolated system is a conserved quantity, the final result of most energy transformations is waste heat transfer to the environment, which is no longer useful for doing work. As we will discuss in more detail in [Thermodynamics](#), the potential for energy to produce useful work has been “degraded” in the energy transformation.

Section Summary

- Power is the rate at which work is done, or in equation form, for the average power P for work W done over a time t , $P = W/t$.
- The SI unit for power is the watt (W), where $1 \text{ W} = 1 \text{ J/s}$.
- The power of many devices such as electric motors is also often expressed in horsepower (hp), where $1 \text{ hp} = 746 \text{ W}$.

Conceptual Questions

Exercise:

Problem:

Most electrical appliances are rated in watts. Does this rating depend on how long the appliance is on? (When off, it is a zero-watt device.) Explain in terms of the definition of power.

Exercise:

Problem:

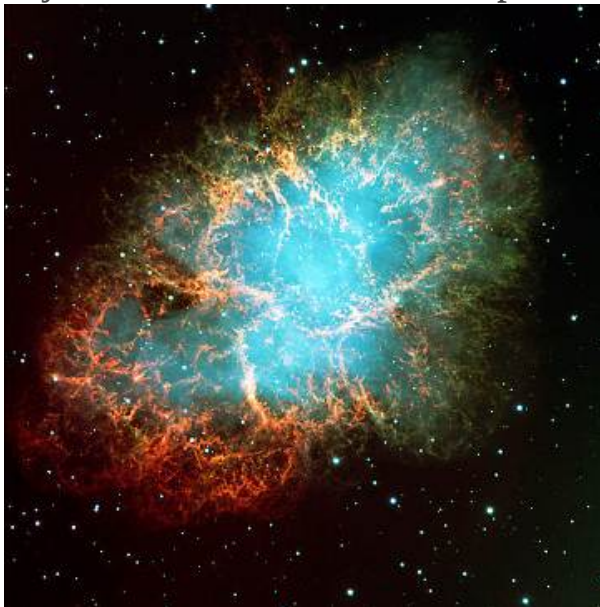
Explain, in terms of the definition of power, why energy consumption is sometimes listed in kilowatt-hours rather than joules. What is the relationship between these two energy units?

Exercise:**Problem:**

A spark of static electricity, such as that you might receive from a doorknob on a cold dry day, may carry a few hundred watts of power. Explain why you are not injured by such a spark.

Problems & Exercises**Exercise:****Problem:**

The Crab Nebula (see [\[link\]](#)) pulsar is the remnant of a supernova that occurred in A.D. 1054. Using data from [\[link\]](#), calculate the approximate factor by which the power output of this astronomical object has declined since its explosion.



Crab Nebula (credit: ESO, via
Wikimedia Commons)

Solution:
Equation:

$$2 \times 10^{-10}$$

Exercise:

Problem:

Suppose a star 1000 times brighter than our Sun (that is, emitting 1000 times the power) suddenly goes supernova. Using data from [\[link\]](#): (a) By what factor does its power output increase? (b) How many times brighter than our entire Milky Way galaxy is the supernova? (c) Based on your answers, discuss whether it should be possible to observe supernovas in distant galaxies. Note that there are on the order of 10^{11} observable galaxies, the average brightness of which is somewhat less than our own galaxy.

Exercise:

Problem:

A person in good physical condition can put out 100 W of useful power for several hours at a stretch, perhaps by pedaling a mechanism that drives an electric generator. Neglecting any problems of generator efficiency and practical considerations such as resting time: (a) How many people would it take to run a 4.00-kW electric clothes dryer? (b) How many people would it take to replace a large electric power plant that generates 800 MW?

Solution:

(a) 40

(b) 8 million

Exercise:

Problem:

What is the cost of operating a 3.00-W electric clock for a year if the cost of electricity is \$0.0900 per $\text{kW} \cdot \text{h}$?

Exercise:

Problem:

A large household air conditioner may consume 15.0 kW of power. What is the cost of operating this air conditioner 3.00 h per day for 30.0 d if the cost of electricity is \$0.110 per $\text{kW} \cdot \text{h}$?

Solution:

\$149

Exercise:

Problem:

(a) What is the average power consumption in watts of an appliance that uses 5.00 $\text{kW} \cdot \text{h}$ of energy per day? (b) How many joules of energy does this appliance consume in a year?

Exercise:

Problem:

(a) What is the average useful power output of a person who does $6.00 \times 10^6 \text{ J}$ of useful work in 8.00 h? (b) Working at this rate, how long will it take this person to lift 2000 kg of bricks 1.50 m to a platform? (Work done to lift his body can be omitted because it is not considered useful output here.)

Solution:

(a) 208 W

(b) 141 s

Exercise:

Problem:

A 500-kg dragster accelerates from rest to a final speed of 110 m/s in 400 m (about a quarter of a mile) and encounters an average frictional force of 1200 N. What is its average power output in watts and horsepower if this takes 7.30 s?

Exercise:

Problem:

(a) How long will it take an 850-kg car with a useful power output of 40.0 hp (1 hp = 746 W) to reach a speed of 15.0 m/s, neglecting friction? (b) How long will this acceleration take if the car also climbs a 3.00-m-high hill in the process?

Solution:

(a) 3.20 s

(b) 4.04 s

Exercise:

Problem:

(a) Find the useful power output of an elevator motor that lifts a 2500-kg load a height of 35.0 m in 12.0 s, if it also increases the speed from rest to 4.00 m/s. Note that the total mass of the counterbalanced system is 10,000 kg—so that only 2500 kg is raised in height, but the full 10,000 kg is accelerated. (b) What does it cost, if electricity is \$0.0900 per kW · h?

Exercise:

Problem:

(a) What is the available energy content, in joules, of a battery that operates a 2.00-W electric clock for 18 months? (b) How long can a battery that can supply 8.00×10^4 J run a pocket calculator that consumes energy at the rate of 1.00×10^{-3} W?

Solution:

(a) 9.46×10^7 J

(b) 2.54 y

Exercise:**Problem:**

(a) How long would it take a 1.50×10^5 -kg airplane with engines that produce 100 MW of power to reach a speed of 250 m/s and an altitude of 12.0 km if air resistance were negligible? (b) If it actually takes 900 s, what is the power? (c) Given this power, what is the average force of air resistance if the airplane takes 1200 s? (Hint: You must find the distance the plane travels in 1200 s assuming constant acceleration.)

Exercise:**Problem:**

Calculate the power output needed for a 950-kg car to climb a 2.00° slope at a constant 30.0 m/s while encountering wind resistance and friction totaling 600 N. Explicitly show how you follow the steps in the [Problem-Solving Strategies for Energy](#).

Solution:

Identify knowns: $m = 950$ kg, slope angle $\theta = 2.00^\circ$, $v = 30.0$ m/s, $f = 600$ N

Identify unknowns: power P of the car, force F that car applies to road

Solve for unknown:

$$P = \frac{W}{t} = \frac{Fd}{t} = F\left(\frac{d}{t}\right) = Fv,$$

where F is parallel to the incline and must oppose the resistive forces and the force of gravity:

$$F = f + w = 600 \text{ N} + mg \sin \theta$$

Insert this into the expression for power and solve:

$$\begin{aligned} P &= (f + mg \sin \theta)v \\ &= \left[600 \text{ N} + (950 \text{ kg}) \left(9.80 \text{ m/s}^2 \right) \sin 2^\circ \right] (30.0 \text{ m/s}) \\ &= 2.77 \times 10^4 \text{ W} \end{aligned}$$

About 28 kW (or about 37 hp) is reasonable for a car to climb a gentle incline.

Exercise:

Problem:

(a) Calculate the power per square meter reaching Earth's upper atmosphere from the Sun. (Take the power output of the Sun to be $4.00 \times 10^{26} \text{ W}$.) (b) Part of this is absorbed and reflected by the atmosphere, so that a maximum of 1.30 kW/m^2 reaches Earth's surface. Calculate the area in km^2 of solar energy collectors needed to replace an electric power plant that generates 750 MW if the collectors convert an average of 2.00% of the maximum power into electricity. (This small conversion efficiency is due to the devices themselves, and the fact that the sun is directly overhead only briefly.) With the same assumptions, what area would be needed to meet the United States' energy needs ($1.05 \times 10^{20} \text{ J}$)? Australia's energy needs ($5.4 \times 10^{18} \text{ J}$)? China's energy needs ($6.3 \times 10^{19} \text{ J}$)? (These energy consumption values are from 2006.)

Glossary

power

the rate at which work is done

watt

(W) SI unit of power, with $1 \text{ W} = 1 \text{ J/s}$

horsepower

an older non-SI unit of power, with $1 \text{ hp} = 746 \text{ W}$

kilowatt-hour

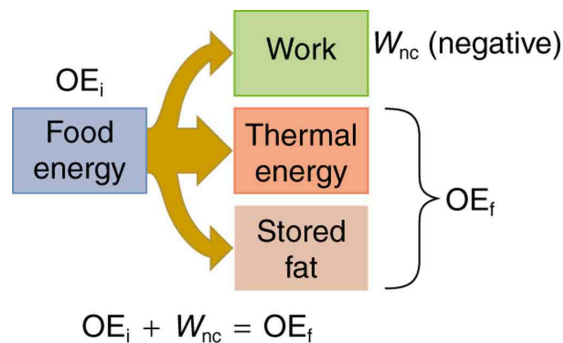
(kW · h) unit used primarily for electrical energy provided by electric utility companies

Work, Energy, and Power in Humans

- Explain the human body's consumption of energy when at rest vs. when engaged in activities that do useful work.
- Calculate the conversion of chemical energy in food into useful work.

Energy Conversion in Humans

Our own bodies, like all living organisms, are energy conversion machines. Conservation of energy implies that the chemical energy stored in food is converted into work, thermal energy, and/or stored as chemical energy in fatty tissue. (See [\[link\]](#).) The fraction going into each form depends both on how much we eat and on our level of physical activity. If we eat more than is needed to do work and stay warm, the remainder goes into body fat.



Energy consumed by humans is converted to work, thermal energy, and stored fat. By far the largest fraction goes to thermal energy, although the fraction varies depending on the type of physical activity.

Power Consumed at Rest

The *rate* at which the body uses food energy to sustain life and to do different activities is called the **metabolic rate**. The total energy conversion rate of a person *at rest* is called the **basal metabolic rate** (BMR) and is divided among various systems in the body, as shown in [\[link\]](#). The largest fraction goes to the liver and spleen, with the brain coming next. Of course, during vigorous exercise, the energy consumption of the skeletal muscles and heart increase markedly. About 75% of the calories burned in a day go into these basic functions. The BMR is a function of age, gender, total body weight, and amount of muscle mass (which burns more calories than body fat). Athletes have a greater BMR due to this last factor.

Organ	Power consumed at rest (W)	Oxygen consumption (mL/min)	Percent of BMR
Liver & spleen	23	67	27
Brain	16	47	19
Skeletal muscle	15	45	18
Kidney	9	26	10
Heart	6	17	7
Other	16	48	19
Totals	85 W	250 mL/min	100%

Basal Metabolic Rates (BMR)

Energy consumption is directly proportional to oxygen consumption because the digestive process is basically one of oxidizing food. We can measure the energy people use during various activities by measuring their oxygen use. (See [\[link\]](#).) Approximately 20 kJ of energy are produced for each liter of oxygen consumed, independent of the type of food. [\[link\]](#) shows energy and oxygen consumption rates (power expended) for a variety of activities.

Power of Doing Useful Work

Work done by a person is sometimes called **useful work**, which is *work done on the outside world*, such as lifting weights. Useful work requires a force exerted through a distance on the outside world, and so it excludes internal work, such as that done by the heart when pumping blood. Useful work does include that done in climbing stairs or accelerating to a full run, because these are accomplished by exerting forces on the outside world. Forces exerted by the body are nonconservative, so that they can change the mechanical energy ($KE + PE$) of the system worked upon, and this is often the goal. A baseball player throwing a ball, for example, increases both the ball's kinetic and potential energy.

If a person needs more energy than they consume, such as when doing vigorous work, the body must draw upon the chemical energy stored in fat. So exercise can be helpful in losing fat. However, the amount of exercise needed to produce a loss in fat, or to burn off extra calories consumed that day, can be large, as [\[link\]](#) illustrates.

Example:

Calculating Weight Loss from Exercising

If a person who normally requires an average of 12,000 kJ (3000 kcal) of food energy per day consumes 13,000 kJ per day, he will steadily gain weight. How much bicycling per day is required to work off this extra 1000 kJ?

Solution

[\[link\]](#) states that 400 W are used when cycling at a moderate speed. The time required to work off 1000 kJ at this rate is then

Equation:

$$\text{Time} = \frac{\text{energy}}{\left(\frac{\text{energy}}{\text{time}}\right)} = \frac{1000 \text{ kJ}}{400 \text{ W}} = 2500 \text{ s} = 42 \text{ min.}$$

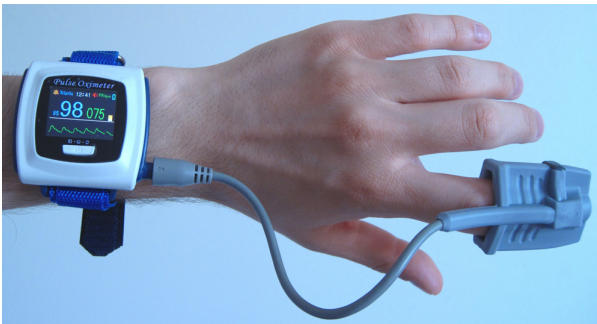
Discussion

If this person uses more energy than he or she consumes, the person's body will obtain the needed energy by metabolizing body fat. If the person uses 13,000 kJ but consumes only 12,000 kJ, then the amount of fat loss will be

Equation:

$$\text{Fat loss} = (1000 \text{ kJ}) \left(\frac{1.0 \text{ g fat}}{39 \text{ kJ}} \right) = 26 \text{ g,}$$

assuming the energy content of fat to be 39 kJ/g.



A pulse oxymeter is an apparatus that measures the amount of oxygen in blood.

Oxymeters can be used to determine a person's metabolic rate, which is the rate at which food energy is converted to another form. Such

measurements can indicate the level of athletic conditioning as well as certain medical problems. (credit: UusiAjaja, Wikimedia Commons)

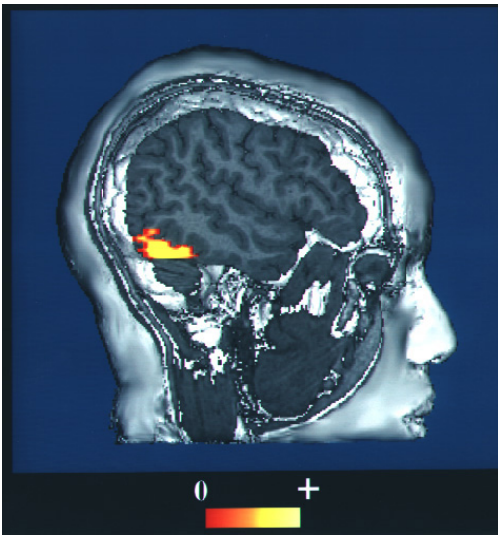
Activity	Energy consumption in watts	Oxygen consumption in liters O₂/min
Sleeping	83	0.24
Sitting at rest	120	0.34
Standing relaxed	125	0.36
Sitting in class	210	0.60
Walking (5 km/h)	280	0.80
Cycling (13–18 km/h)	400	1.14
Shivering	425	1.21
Playing tennis	440	1.26

Activity	Energy consumption in watts	Oxygen consumption in liters O₂/min
Swimming breaststroke	475	1.36
Ice skating (14.5 km/h)	545	1.56
Climbing stairs (116/min)	685	1.96
Cycling (21 km/h)	700	2.00
Running cross-country	740	2.12
Playing basketball	800	2.28
Cycling, professional racer	1855	5.30
Sprinting	2415	6.90

Energy and Oxygen Consumption Rates[\[footnote\]](#) (Power)
for an average 76-kg male

All bodily functions, from thinking to lifting weights, require energy. (See [\[link\]](#).) The many small muscle actions accompanying all quiet activity, from sleeping to head scratching, ultimately become thermal energy, as do less visible muscle actions by the heart, lungs, and digestive tract. Shivering, in fact, is an involuntary response to low body temperature that pits muscles against one another to produce thermal energy in the body (and

do no work). The kidneys and liver consume a surprising amount of energy, but the biggest surprise of all is that a full 25% of all energy consumed by the body is used to maintain electrical potentials in all living cells. (Nerve cells use this electrical potential in nerve impulses.) This bioelectrical energy ultimately becomes mostly thermal energy, but some is utilized to power chemical processes such as in the kidneys and liver, and in fat production.



This fMRI scan shows an increased level of energy consumption in the vision center of the brain. Here, the patient was being asked to recognize faces.
(credit: NIH via Wikimedia Commons)

Section Summary

- The human body converts energy stored in food into work, thermal energy, and/or chemical energy that is stored in fatty tissue.
- The *rate* at which the body uses food energy to sustain life and to do different activities is called the metabolic rate, and the corresponding rate when at rest is called the basal metabolic rate (BMR)
- The energy included in the basal metabolic rate is divided among various systems in the body, with the largest fraction going to the liver and spleen, and the brain coming next.
- About 75% of food calories are used to sustain basic body functions included in the basal metabolic rate.
- The energy consumption of people during various activities can be determined by measuring their oxygen use, because the digestive process is basically one of oxidizing food.

Conceptual Questions

Exercise:

Problem:

Explain why it is easier to climb a mountain on a zigzag path rather than one straight up the side. Is your increase in gravitational potential energy the same in both cases? Is your energy consumption the same in both?

Exercise:

Problem:

Do you do work on the outside world when you rub your hands together to warm them? What is the efficiency of this activity?

Exercise:

Problem:

Shivering is an involuntary response to lowered body temperature. What is the efficiency of the body when shivering, and is this a desirable value?

Exercise:

Problem:

Discuss the relative effectiveness of dieting and exercise in losing weight, noting that most athletic activities consume food energy at a rate of 400 to 500 W, while a single cup of yogurt can contain 1360 kJ (325 kcal). Specifically, is it likely that exercise alone will be sufficient to lose weight? You may wish to consider that regular exercise may increase the metabolic rate, whereas protracted dieting may reduce it.

Problems & Exercises**Exercise:****Problem:**

(a) How long can you rapidly climb stairs (116/min) on the 93.0 kcal of energy in a 10.0-g pat of butter? (b) How many flights is this if each flight has 16 stairs?

Solution:

(a) 9.5 min

(b) 69 flights of stairs

Exercise:**Problem:**

(a) What is the power output in watts and horsepower of a 70.0-kg sprinter who accelerates from rest to 10.0 m/s in 3.00 s? (b) Considering the amount of power generated, do you think a well-trained athlete could do this repetitively for long periods of time?

Exercise:

Problem:

Calculate the power output in watts and horsepower of a shot-putter who takes 1.20 s to accelerate the 7.27-kg shot from rest to 14.0 m/s, while raising it 0.800 m. (Do not include the power produced to accelerate his body.)



Shot putter at the
Dornoch Highland
Gathering in 2007.
(credit: John Haslam,
Flickr)

Solution:

641 W, 0.860 hp

Exercise:**Problem:**

(a) What is the efficiency of an out-of-condition professor who does 2.10×10^5 J of useful work while metabolizing 500 kcal of food energy? (b) How many food calories would a well-conditioned athlete metabolize in doing the same work with an efficiency of 20%?

Exercise:

Problem:

Energy that is not utilized for work or heat transfer is converted to the chemical energy of body fat containing about 39 kJ/g. How many grams of fat will you gain if you eat 10,000 kJ (about 2500 kcal) one day and do nothing but sit relaxed for 16.0 h and sleep for the other 8.00 h? Use data from [\[link\]](#) for the energy consumption rates of these activities.

Solution:

31 g

Exercise:**Problem:**

Using data from [\[link\]](#), calculate the daily energy needs of a person who sleeps for 7.00 h, walks for 2.00 h, attends classes for 4.00 h, cycles for 2.00 h, sits relaxed for 3.00 h, and studies for 6.00 h. (Studying consumes energy at the same rate as sitting in class.)

Exercise:**Problem:**

What is the efficiency of a subject on a treadmill who puts out work at the rate of 100 W while consuming oxygen at the rate of 2.00 L/min? (Hint: See [\[link\]](#).)

Solution:

14.3%

Exercise:

Problem:

Shoveling snow can be extremely taxing because the arms have such a low efficiency in this activity. Suppose a person shoveling a footpath metabolizes food at the rate of 800 W. (a) What is her useful power output? (b) How long will it take her to lift 3000 kg of snow 1.20 m? (This could be the amount of heavy snow on 20 m of footpath.) (c) How much waste heat transfer in kilojoules will she generate in the process?

Exercise:**Problem:**

Very large forces are produced in joints when a person jumps from some height to the ground. (a) Calculate the magnitude of the force produced if an 80.0-kg person jumps from a 0.600-m-high ledge and lands stiffly, compressing joint material 1.50 cm as a result. (Be certain to include the weight of the person.) (b) In practice the knees bend almost involuntarily to help extend the distance over which you stop. Calculate the magnitude of the force produced if the stopping distance is 0.300 m. (c) Compare both forces with the weight of the person.

Solution:

(a) $3.21 \times 10^4 \text{ N}$

(b) $2.35 \times 10^3 \text{ N}$

(c) Ratio of net force to weight of person is 41.0 in part (a); 3.00 in part (b)

Exercise:

Problem:

Jogging on hard surfaces with insufficiently padded shoes produces large forces in the feet and legs. (a) Calculate the magnitude of the force needed to stop the downward motion of a jogger's leg, if his leg has a mass of 13.0 kg, a speed of 6.00 m/s, and stops in a distance of 1.50 cm. (Be certain to include the weight of the 75.0-kg jogger's body.) (b) Compare this force with the weight of the jogger.

Exercise:**Problem:**

(a) Calculate the energy in kJ used by a 55.0-kg woman who does 50 deep knee bends in which her center of mass is lowered and raised 0.400 m. (She does work in both directions.) You may assume her efficiency is 20%. (b) What is the average power consumption rate in watts if she does this in 3.00 min?

Solution:

(a) 108 kJ

(b) 599 W

Exercise:**Problem:**

Kanellos Kanellopoulos flew 119 km from Crete to Santorini, Greece, on April 23, 1988, in the *Daedalus 88*, an aircraft powered by a bicycle-type drive mechanism (see [\[link\]](#)). His useful power output for the 234-min trip was about 350 W. Using the efficiency for cycling from [\[link\]](#), calculate the food energy in kilojoules he metabolized during the flight.

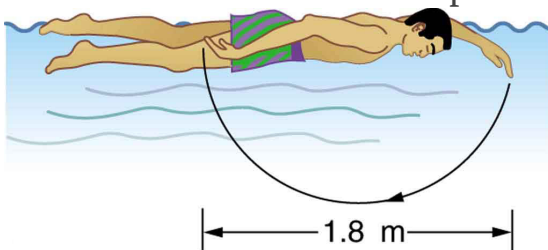


The Daedalus 88 in flight.
(credit: NASA photo by
Beasley)

Exercise:

Problem:

The swimmer shown in [\[link\]](#) exerts an average horizontal backward force of 80.0 N with his arm during each 1.80 m long stroke. (a) What is his work output in each stroke? (b) Calculate the power output of his arms if he does 120 strokes per minute.



Solution:

(a) 144 J

(b) 288 W

Exercise:

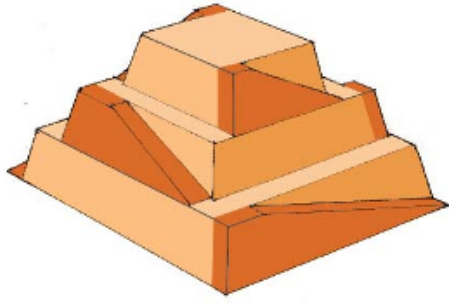
Problem:

Mountain climbers carry bottled oxygen when at very high altitudes.

(a) Assuming that a mountain climber uses oxygen at twice the rate for climbing 116 stairs per minute (because of low air temperature and winds), calculate how many liters of oxygen a climber would need for 10.0 h of climbing. (These are liters at sea level.) Note that only 40% of the inhaled oxygen is utilized; the rest is exhaled. (b) How much useful work does the climber do if he and his equipment have a mass of 90.0 kg and he gains 1000 m of altitude? (c) What is his efficiency for the 10.0-h climb?

Exercise:**Problem:**

The awe-inspiring Great Pyramid of Cheops was built more than 4500 years ago. Its square base, originally 230 m on a side, covered 13.1 acres, and it was 146 m high, with a mass of about 7×10^9 kg. (The pyramid's dimensions are slightly different today due to quarrying and some sagging.) Historians estimate that 20,000 workers spent 20 years to construct it, working 12-hour days, 330 days per year. (a) Calculate the gravitational potential energy stored in the pyramid, given its center of mass is at one-fourth its height. (b) Only a fraction of the workers lifted blocks; most were involved in support services such as building ramps (see [\[link\]](#)), bringing food and water, and hauling blocks to the site. Calculate the efficiency of the workers who did the lifting, assuming there were 1000 of them and they consumed food energy at the rate of 300 kcal/h. What does your answer imply about how much of their work went into block-lifting, versus how much work went into friction and lifting and lowering their own bodies? (c) Calculate the mass of food that had to be supplied each day, assuming that the average worker required 3600 kcal per day and that their diet was 5% protein, 60% carbohydrate, and 35% fat. (These proportions neglect the mass of bulk and nondigestible materials consumed.)



Ancient pyramids were probably constructed using ramps as simple machines.
(credit: Franck Monnier, Wikimedia Commons)

Solution:

- (a) $2.50 \times 10^{12} \text{ J}$
- (b) 2.52%
- (c) $1.4 \times 10^4 \text{ kg}$ (14 metric tons)

Exercise:

Problem:

(a) How long can you play tennis on the 800 kJ (about 200 kcal) of energy in a candy bar? (b) Does this seem like a long time? Discuss why exercise is necessary but may not be sufficient to cause a person to lose weight.

Glossary

metabolic rate

the rate at which the body uses food energy to sustain life and to do different activities

basal metabolic rate

the total energy conversion rate of a person at rest

useful work

work done on an external system

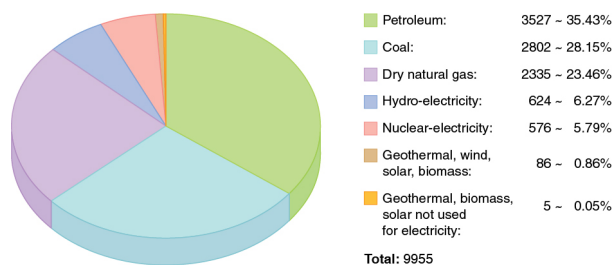
World Energy Use

- Describe the distinction between renewable and nonrenewable energy sources.
- Explain why the inevitable conversion of energy to less useful forms makes it necessary to conserve energy resources.

Energy is an important ingredient in all phases of society. We live in a very interdependent world, and access to adequate and reliable energy resources is crucial for economic growth and for maintaining the quality of our lives. But current levels of energy consumption and production are not sustainable. About 40% of the world's energy comes from oil, and much of that goes to transportation uses. Oil prices are dependent as much upon new (or foreseen) discoveries as they are upon political events and situations around the world. The U.S., with 4.5% of the world's population, consumes 24% of the world's oil production per year; 66% of that oil is imported!

Renewable and Nonrenewable Energy Sources

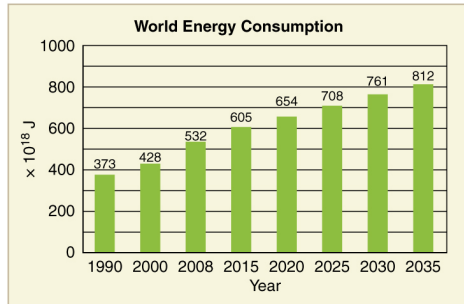
The principal energy resources used in the world are shown in [\[link\]](#). The fuel mix has changed over the years but now is dominated by oil, although natural gas and solar contributions are increasing. **Renewable forms of energy** are those sources that cannot be used up, such as water, wind, solar, and biomass. About 85% of our energy comes from nonrenewable **fossil fuels**—oil, natural gas, coal. The likelihood of a link between global warming and fossil fuel use, with its production of carbon dioxide through combustion, has made, in the eyes of many scientists, a shift to non-fossil fuels of utmost importance—but it will not be easy.



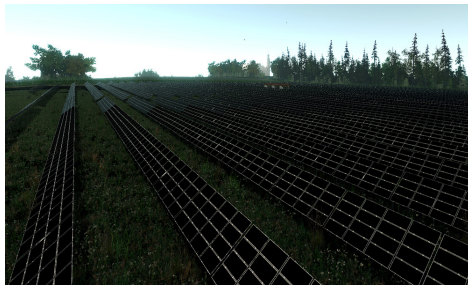
World energy consumption by source, in billions of kilowatt-hours: 2006. (credit: KVDP)

The World's Growing Energy Needs

World energy consumption continues to rise, especially in the developing countries. (See [\[link\]](#).) Global demand for energy has tripled in the past 50 years and might triple again in the next 30 years. While much of this growth will come from the rapidly booming economies of China and India, many of the developed countries, especially those in Europe, are hoping to meet their energy needs by expanding the use of renewable sources. Although presently only a small percentage, renewable energy is growing very fast, especially wind energy. For example, Germany plans to meet 20% of its electricity and 10% of its overall energy needs with renewable resources by the year 2020. (See [\[link\]](#).) Energy is a key constraint in the rapid economic growth of China and India. In 2003, China surpassed Japan as the world's second largest consumer of oil. However, over 1/3 of this is imported. Unlike most Western countries, coal dominates the commercial energy resources of China, accounting for 2/3 of its energy consumption. In 2009 China surpassed the United States as the largest generator of CO₂. In India, the main energy resources are biomass (wood and dung) and coal. Half of India's oil is imported. About 70% of India's electricity is generated by highly polluting coal. Yet there are sizeable strides being made in renewable energy. India has a rapidly growing wind energy base, and it has the largest solar cooking program in the world.



Past and projected world energy use
(source: Based on data from U.S.
Energy Information Administration,
2011)



Solar cell arrays at a power plant in
Steindorf, Germany (credit: Michael
Betke, Flickr)

[\[link\]](#) displays the 2006 commercial energy mix by country for some of the prime energy users in the world. While non-renewable sources dominate, some countries get a sizeable percentage of their electricity from renewable resources. For example, about 67% of New Zealand's electricity demand is met by hydroelectric. Only 10% of the U.S. electricity is generated by renewable resources, primarily hydroelectric. It is difficult to determine total contributions of renewable energy in some countries with a large rural population, so these percentages in this table are left blank.

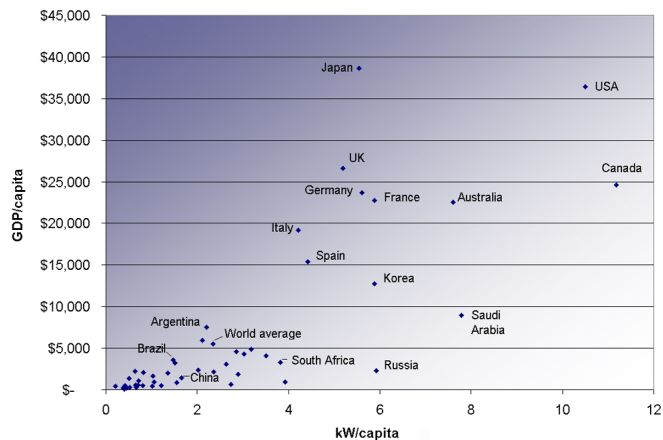
Country	Consumption, in EJ (10 ¹⁸ J)	Oil	Natural Gas	Coal	Nuclear	Hydro	Other Renewables
Australia	5.4	34%	17%	44%	0%	3%	1%

Country	Consumption, in EJ (10^{18} J)	Oil	Natural Gas	Coal	Nuclear	Hydro	Other Renewables
Brazil	9.6	48%	7%	5%	1%	35%	2%
China	63	22%	3%	69%	1%	6%	
Egypt	2.4	50%	41%	1%	0%	6%	
Germany	16	37%	24%	24%	11%	1%	3%
India	15	34%	7%	52%	1%	5%	
Indonesia	4.9	51%	26%	16%	0%	2%	3%
Japan	24	48%	14%	21%	12%	4%	1%
New Zealand	0.44	32%	26%	6%	0%	11%	19%
Russia	31	19%	53%	16%	5%	6%	
U.S.	105	40%	23%	22%	8%	3%	1%
World	432	39%	23%	24%	6%	6%	2%

Energy Consumption—Selected Countries (2006)

Energy and Economic Well-being

The last two columns in this table examine the energy and electricity use per capita. Economic well-being is dependent upon energy use, and in most countries higher standards of living, as measured by GDP (gross domestic product) per capita, are matched by higher levels of energy consumption per capita. This is borne out in [\[link\]](#). Increased efficiency of energy use will change this dependency. A global problem is balancing energy resource development against the harmful effects upon the environment in its extraction and use.



Power consumption per capita versus GDP per capita for various countries. Note the increase in energy usage with increasing GDP. (2007, credit: Frank van Mierlo, Wikimedia Commons)

Conserving Energy

As we finish this chapter on energy and work, it is relevant to draw some distinctions between two sometimes misunderstood terms in the area of energy use. As has been mentioned elsewhere, the “law of the conservation of energy” is a very useful principle in analyzing physical processes. It is a statement that cannot be proven from basic principles, but is a very good bookkeeping device, and no exceptions have ever been found. It states that the total amount of energy in an isolated system will always remain constant. Related to this principle, but remarkably different from it, is the important philosophy of energy conservation. This concept has to do with seeking to decrease the amount of energy used by an individual or group through (1) reduced activities (e.g., turning down thermostats, driving fewer kilometers) and/or (2) increasing conversion efficiencies in the performance of a particular task—such as developing and using more efficient room heaters, cars that have greater miles-per-gallon ratings, energy-efficient compact fluorescent lights, etc.

Since energy in an isolated system is not destroyed or created or generated, one might wonder why we need to be concerned about our energy resources, since energy is a conserved quantity. The problem is that the final result of most energy transformations is waste heat transfer to the environment and conversion to energy forms no longer useful for doing work. To state it in another way, the potential for energy to produce useful work has been “degraded” in the energy transformation. (This will be discussed in more detail in [Thermodynamics](#).)

Section Summary

- The relative use of different fuels to provide energy has changed over the years, but fuel use is currently dominated by oil, although natural gas and solar contributions are increasing.
- Although non-renewable sources dominate, some countries meet a sizeable percentage of their electricity needs from renewable resources.
- The United States obtains only about 10% of its energy from renewable sources, mostly hydroelectric power.
- Economic well-being is dependent upon energy use, and in most countries higher standards of living, as measured by GDP (Gross Domestic Product) per capita, are matched by higher levels of energy consumption per capita.
- Even though, in accordance with the law of conservation of energy, energy can never be created or destroyed, energy that can be used to do work is always partly converted to less useful forms, such as waste heat to the environment, in all of our uses of energy for practical purposes.

Conceptual Questions

Exercise:

Problem:

What is the difference between energy conservation and the law of conservation of energy? Give some examples of each.

Exercise:

Problem:

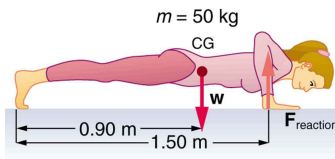
If the efficiency of a coal-fired electrical generating plant is 35%, then what do we mean when we say that energy is a conserved quantity?

Problems & Exercises

Exercise:

Problem: Integrated Concepts

(a) Calculate the force the woman in [\[link\]](#) exerts to do a push-up at constant speed, taking all data to be known to three digits. (b) How much work does she do if her center of mass rises 0.240 m? (c) What is her useful power output if she does 25 push-ups in 1 min? (Should work done lowering her body be included? See the discussion of useful work in [Work, Energy, and Power in Humans](#).



Forces involved in doing push-ups. The woman's weight acts as a force exerted downward on her center of gravity (CG).

Solution:

- (a) 294 N
- (b) 118 J
- (c) 49.0 W

Exercise:

Problem: Integrated Concepts

A 75.0-kg cross-country skier is climbing a 3.0° slope at a constant speed of 2.00 m/s and encounters air resistance of 25.0 N. Find his power output for work done against the gravitational force and air resistance. (b) What average force does he exert backward on the snow to accomplish this? (c) If he continues to exert this force and to experience the same air resistance when he reaches a level area, how long will it take him to reach a velocity of 10.0 m/s?

Exercise:**Problem: Integrated Concepts**

The 70.0-kg swimmer in [\[link\]](#) starts a race with an initial velocity of 1.25 m/s and exerts an average force of 80.0 N backward with his arms during each 1.80 m long stroke. (a) What is his initial acceleration if water resistance is 45.0 N? (b) What is the subsequent average resistance force from the water during the 5.00 s it takes him to reach his top velocity of 2.50 m/s? (c) Discuss whether water resistance seems to increase linearly with velocity.

Solution:

(a) 0.500 m/s^2

(b) 62.5 N

(c) Assuming the acceleration of the swimmer decreases linearly with time over the 5.00 s interval, the frictional force must therefore be increasing linearly with time, since $f = F - ma$. If the acceleration decreases linearly with time, the velocity will contain a term dependent on time squared (t^2). Therefore, the water resistance will not depend linearly on the velocity.

Exercise:**Problem: Integrated Concepts**

A toy gun uses a spring with a force constant of 300 N/m to propel a 10.0-g steel ball. If the spring is compressed 7.00 cm and friction is negligible: (a) How much force is needed to compress the spring? (b) To what maximum height can the ball be shot? (c) At what angles above the horizontal may a child aim to hit a target 3.00 m away at the same height as the gun? (d) What is the gun's maximum range on level ground?

Exercise:**Problem: Integrated Concepts**

(a) What force must be supplied by an elevator cable to produce an acceleration of 0.800 m/s^2 against a 200-N frictional force, if the mass of the loaded elevator is 1500 kg? (b) How much work is done by the cable in lifting the elevator 20.0 m? (c) What is the final speed of the elevator if it starts from rest? (d) How much work went into thermal energy?

Solution:

(a) $16.1 \times 10^3 \text{ N}$

(b) $3.22 \times 10^5 \text{ J}$

(c) 5.66 m/s

(d) 4.00 kJ

Exercise:**Problem: Unreasonable Results**

A car advertisement claims that its 900-kg car accelerated from rest to 30.0 m/s and drove 100 km, gaining 3.00 km in altitude, on 1.0 gal of gasoline. The average force of friction including air resistance was 700 N. Assume all values are known to three significant figures. (a) Calculate the car's efficiency. (b) What is unreasonable about the result? (c) Which premise is unreasonable, or which premises are inconsistent?

Exercise:**Problem: Unreasonable Results**

Body fat is metabolized, supplying 9.30 kcal/g, when dietary intake is less than needed to fuel metabolism. The manufacturers of an exercise bicycle claim that you can lose 0.500 kg of fat per day by vigorously exercising for 2.00 h per day on their machine. (a) How many kcal are supplied by the metabolization of 0.500 kg of fat? (b) Calculate the kcal/min that you would have to utilize to metabolize fat at the rate of 0.500 kg in 2.00 h. (c) What is unreasonable about the results? (d) Which premise is unreasonable, or which premises are inconsistent?

Solution:

(a) 4.65×10^3 kcal

(b) 38.8 kcal/min

(c) This power output is higher than the highest value on [\[link\]](#), which is about 35 kcal/min (corresponding to 2415 watts) for sprinting.

(d) It would be impossible to maintain this power output for 2 hours (imagine sprinting for 2 hours!).

Exercise:**Problem: Construct Your Own Problem**

Consider a person climbing and descending stairs. Construct a problem in which you calculate the long-term rate at which stairs can be climbed considering the mass of the person, his ability to generate power with his legs, and the height of a single stair step. Also consider why the same person can descend stairs at a faster rate for a nearly unlimited time in spite of the fact that very similar forces are exerted going down as going up. (This points to a fundamentally different process for descending versus climbing stairs.)

Exercise:**Problem: Construct Your Own Problem**

Consider humans generating electricity by pedaling a device similar to a stationary bicycle. Construct a problem in which you determine the number of people it would take to replace a large electrical generation facility. Among the things to consider are the power output that is reasonable using the legs, rest time, and the need for electricity 24 hours per day. Discuss the practical implications of your results.

Exercise:**Problem: Integrated Concepts**

A 105-kg basketball player crouches down 0.400 m while waiting to jump. After exerting a force on the floor through this 0.400 m, his feet leave the floor and his center of gravity rises 0.950 m above its normal standing erect position. (a) Using energy considerations, calculate his velocity when he leaves the floor. (b) What average force did he exert on the floor? (Do not neglect the force to support his weight as well as that to accelerate him.) (c) What was his power output during the acceleration phase?

Solution:

(a) 4.32 m/s

(b) 3.47×10^3 N

(c) 8.93 kW

Glossary

renewable forms of energy

those sources that cannot be used up, such as water, wind, solar, and biomass

fossil fuels

oil, natural gas, and coal

Temperature

- Define temperature.
- Convert temperatures between the Celsius, Fahrenheit, and Kelvin scales.
- Define thermal equilibrium.
- State the zeroth law of thermodynamics.

The concept of temperature has evolved from the common concepts of hot and cold. Human perception of what feels hot or cold is a relative one. For example, if you place one hand in hot water and the other in cold water, and then place both hands in tepid water, the tepid water will feel cool to the hand that was in hot water, and warm to the one that was in cold water. The scientific definition of temperature is less ambiguous than your senses of hot and cold. **Temperature** is operationally defined to be what we measure with a thermometer. (Many physical quantities are defined solely in terms of how they are measured. We shall see later how temperature is related to the kinetic energies of atoms and molecules, a more physical explanation.) Two accurate thermometers, one placed in hot water and the other in cold water, will show the hot water to have a higher temperature. If they are then placed in the tepid water, both will give identical readings (within measurement uncertainties). In this section, we discuss temperature, its measurement by thermometers, and its relationship to thermal equilibrium. Again, temperature is the quantity measured by a thermometer.

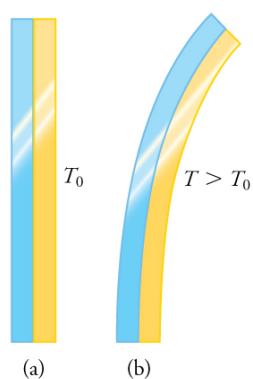
Note:

Misconception Alert: Human Perception vs. Reality

On a cold winter morning, the wood on a porch feels warmer than the metal of your bike. The wood and bicycle are in thermal equilibrium with the outside air, and are thus the same temperature. They *feel* different because of the difference in the way that they conduct heat away from your skin. The metal conducts heat away from your body faster than the wood does (see more about conductivity in [Conduction](#)). This is just one example demonstrating that the human sense of hot and cold is not determined by temperature alone.

Another factor that affects our perception of temperature is humidity. Most people feel much hotter on hot, humid days than on hot, dry days. This is because on humid days, sweat does not evaporate from the skin as efficiently as it does on dry days. It is the evaporation of sweat (or water from a sprinkler or pool) that cools us off.

Any physical property that depends on temperature, and whose response to temperature is reproducible, can be used as the basis of a thermometer. Because many physical properties depend on temperature, the variety of thermometers is remarkable. For example, volume increases with temperature for most substances. This property is the basis for the common alcohol thermometer, the old mercury thermometer, and the bimetallic strip ([link](#)). Other properties used to measure temperature include electrical resistance and color, as shown in [link](#), and the emission of infrared radiation, as shown in [link](#).



The curvature of a bimetallic strip depends on

temperature.
(a) The strip is straight at the starting temperature, where its two components have the same length.

(b) At a higher temperature, this strip bends to the right, because the metal on the left has expanded more than the metal on the right.



Each of the six squares on this plastic (liquid crystal)

thermometer contains a film of a different heat-sensitive liquid crystal material. Below 95°F, all six squares are black.

When the plastic thermometer is exposed to temperature that increases to 95°F, the first liquid crystal square changes color. When the temperature increases above 96.8°F the second liquid crystal square also changes color, and so forth. (credit: Arkrishna, Wikimedia Commons)



Fireman Jason Ormand uses a pyrometer to check the temperature of an aircraft carrier's ventilation system. Infrared radiation (whose emission varies with temperature)

from the vent is measured and a temperature readout is quickly produced.

Infrared measurements are also frequently used as a measure of body temperature.

These modern thermometers, placed in the ear canal, are more accurate than alcohol thermometers placed under the tongue or in the armpit. (credit: Lamel J. Hinton/U.S. Navy)

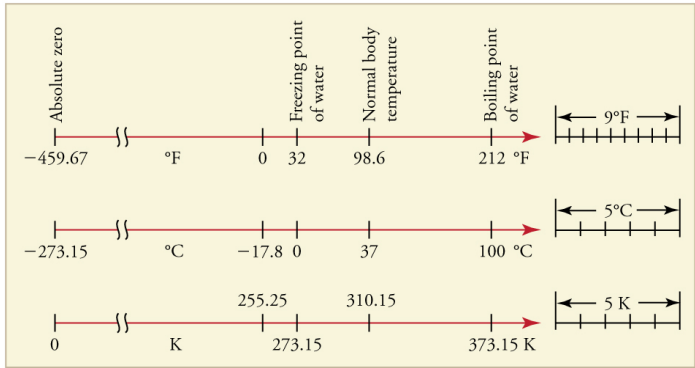
Temperature Scales

Thermometers are used to measure temperature according to well-defined scales of measurement, which use pre-defined reference points to help compare quantities. The three most common temperature scales are the Fahrenheit, Celsius, and Kelvin scales. A temperature scale can be created by identifying two easily reproducible temperatures. The freezing and boiling temperatures of water at standard atmospheric pressure are commonly used.

The **Celsius** scale (which replaced the slightly different *centigrade* scale) has the freezing point of water at 0°C and the boiling point at 100°C . Its unit is the **degree Celsius** ($^{\circ}\text{C}$). On the **Fahrenheit** scale (still the most frequently used in the United States), the freezing point of water is at 32°F and the boiling point is at 212°F . The unit of temperature on this scale is the **degree Fahrenheit** ($^{\circ}\text{F}$). Note that a temperature difference of one degree Celsius is greater than a temperature difference of one degree Fahrenheit. Only 100 Celsius degrees

span the same range as 180 Fahrenheit degrees, thus one degree on the Celsius scale is 1.8 times larger than one degree on the Fahrenheit scale $180/100 = 9/5$.

The **Kelvin** scale is the temperature scale that is commonly used in science. It is an *absolute temperature* scale defined to have 0 K at the lowest possible temperature, called **absolute zero**. The official temperature unit on this scale is the *kelvin*, which is abbreviated K, and is not accompanied by a degree sign. The freezing and boiling points of water are 273.15 K and 373.15 K, respectively. Thus, the magnitude of temperature differences is the same in units of kelvins and degrees Celsius. Unlike other temperature scales, the Kelvin scale is an absolute scale. It is used extensively in scientific work because a number of physical quantities, such as the volume of an ideal gas, are directly related to absolute temperature. The kelvin is the SI unit used in scientific work.



Relationships between the Fahrenheit, Celsius, and Kelvin temperature scales, rounded to the nearest degree. The relative sizes of the scales are also shown.

The relationships between the three common temperature scales is shown in [\[link\]](#). Temperatures on these scales can be converted using the equations in [\[link\]](#).

To convert from ...	Use this equation ...	Also written as ...
---------------------	-----------------------	---------------------

To convert from ...	Use this equation ...	Also written as ...
Celsius to Fahrenheit	$T(^{\circ}\text{F}) = \frac{9}{5}T(^{\circ}\text{C}) + 32$	$T_{\text{F}} = \frac{9}{5}T_{\text{C}} + 32$
Fahrenheit to Celsius	$T(^{\circ}\text{C}) = \frac{5}{9}(T(^{\circ}\text{F}) - 32)$	$T_{\text{C}} = \frac{5}{9}(T_{\text{F}} - 32)$
Celsius to Kelvin	$T(\text{K}) = T(^{\circ}\text{C}) + 273.15$	$T_{\text{K}} = T_{\text{C}} + 273.15$
Kelvin to Celsius	$T(^{\circ}\text{C}) = T(\text{K}) - 273.15$	$T_{\text{C}} = T_{\text{K}} - 273.15$
Fahrenheit to Kelvin	$T(\text{K}) = \frac{5}{9}(T(^{\circ}\text{F}) - 32) + 273.15$	$T_{\text{K}} = \frac{5}{9}(T_{\text{F}} - 32) + 273.15$
Kelvin to Fahrenheit	$T(^{\circ}\text{F}) = \frac{9}{5}(T(\text{K}) - 273.15) + 32$	$T_{\text{F}} = \frac{9}{5}(T_{\text{K}} - 273.15) + 32$

Temperature Conversions

Notice that the conversions between Fahrenheit and Kelvin look quite complicated. In fact, they are simple combinations of the conversions between Fahrenheit and Celsius, and the conversions between Celsius and Kelvin.

Example:

Converting between Temperature Scales: Room Temperature

“Room temperature” is generally defined to be 25°C . (a) What is room temperature in $^{\circ}\text{F}$?

(b) What is it in K?

Strategy

To answer these questions, all we need to do is choose the correct conversion equations and plug in the known values.

Solution for (a)

1. Choose the right equation. To convert from °C to °F, use the equation

Equation:

$$T_{\text{°F}} = \frac{9}{5}T_{\text{°C}} + 32.$$

2. Plug the known value into the equation and solve:

Equation:

$$T_{\text{°F}} = \frac{9}{5}25^{\circ}\text{C} + 32 = 77^{\circ}\text{F}.$$

Solution for (b)

1. Choose the right equation. To convert from °C to K, use the equation

Equation:

$$T_{\text{K}} = T_{\text{°C}} + 273.15.$$

2. Plug the known value into the equation and solve:

Equation:

$$T_{\text{K}} = 25^{\circ}\text{C} + 273.15 = 298 \text{ K}.$$

Example:**Converting between Temperature Scales: the Reaumur Scale**

The Reaumur scale is a temperature scale that was used widely in Europe in the 18th and 19th centuries. On the Reaumur temperature scale, the freezing point of water is 0°R and the boiling temperature is 80°R. If “room temperature” is 25°C on the Celsius scale, what is it on the Reaumur scale?

Strategy

To answer this question, we must compare the Reaumur scale to the Celsius scale. The difference between the freezing point and boiling point of water on the Reaumur scale is 80°R. On the Celsius scale it is 100°C. Therefore $100^{\circ}\text{C} = 80^{\circ}\text{R}$. Both scales start at 0° for freezing, so we can derive a simple formula to convert between temperatures on the two scales.

Solution

1. Derive a formula to convert from one scale to the other:

Equation:

$$T_{\text{°R}} = \frac{0.8^{\circ}\text{R}}{^{\circ}\text{C}} \times T_{\text{°C}}.$$

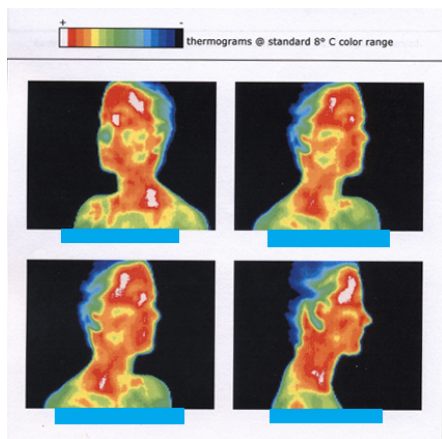
2. Plug the known value into the equation and solve:

Equation:

$$T_{\text{R}} = \frac{0.8^{\circ}\text{R}}{^{\circ}\text{C}} \times 25^{\circ}\text{C} = 20^{\circ}\text{R}.$$

Temperature Ranges in the Universe

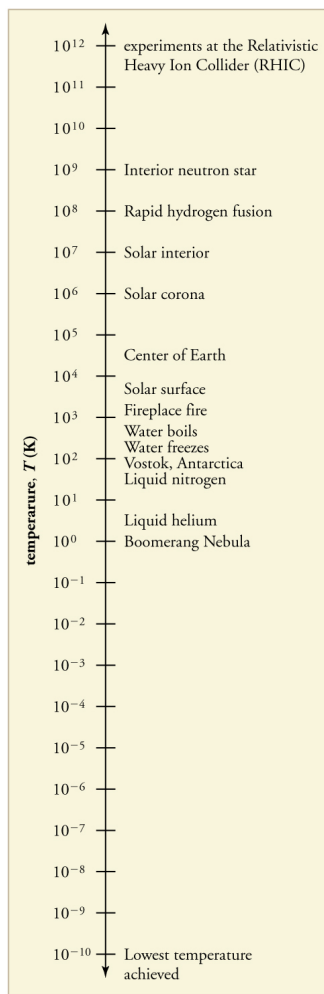
[\[link\]](#) shows the wide range of temperatures found in the universe. Human beings have been known to survive with body temperatures within a small range, from 24°C to 44°C (75°F to 111°F). The average normal body temperature is usually given as 37.0°C (98.6°F), and variations in this temperature can indicate a medical condition: a fever, an infection, a tumor, or circulatory problems (see [\[link\]](#)).



This image of radiation from a person's body (an infrared thermograph) shows the location of temperature abnormalities in the upper body. Dark blue corresponds to cold areas and red to white corresponds to hot areas. An elevated temperature might be an indication of malignant tissue (a cancerous tumor in the breast, for example), while a depressed temperature

might be due to a decline in blood flow from a clot. In this case, the abnormalities are caused by a condition called hyperhidrosis.
(credit: Porcelina81, Wikimedia Commons)

The lowest temperatures ever recorded have been measured during laboratory experiments: 4.5×10^{-10} K at the Massachusetts Institute of Technology (USA), and 1.0×10^{-10} K at Helsinki University of Technology (Finland). In comparison, the coldest recorded place on Earth's surface is Vostok, Antarctica at 183 K (-89°C), and the coldest place (outside the lab) known in the universe is the Boomerang Nebula, with a temperature of 1 K.

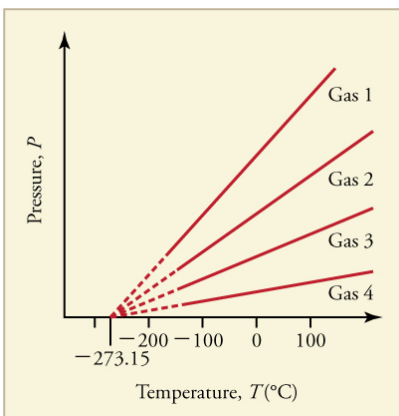


Each increment on this logarithmic scale indicates an increase by a factor of ten, and thus illustrates the tremendous range of temperatures in nature. Note that zero on a logarithmic scale would occur off the bottom of the page at infinity.

Note:

Making Connections: Absolute Zero

What is absolute zero? Absolute zero is the temperature at which all molecular motion has ceased. The concept of absolute zero arises from the behavior of gases. [\[link\]](#) shows how the pressure of gases at a constant volume decreases as temperature decreases. Various scientists have noted that the pressures of gases extrapolate to zero at the same temperature, -273.15°C . This extrapolation implies that there is a lowest temperature. This temperature is called *absolute zero*. Today we know that most gases first liquefy and then freeze, and it is not actually possible to reach absolute zero. The numerical value of absolute zero temperature is -273.15°C or 0 K.



Graph of pressure versus temperature for various

gases kept at a constant volume. Note that all of the graphs extrapolate to zero pressure at the same temperature.

Thermal Equilibrium and the Zeroth Law of Thermodynamics

Thermometers actually take their *own* temperature, not the temperature of the object they are measuring. This raises the question of how we can be certain that a thermometer measures the temperature of the object with which it is in contact. It is based on the fact that any two systems placed in *thermal contact* (meaning heat transfer can occur between them) will reach the same temperature. That is, heat will flow from the hotter object to the cooler one until they have exactly the same temperature. The objects are then in **thermal equilibrium**, and no further changes will occur. The systems interact and change because their temperatures differ, and the changes stop once their temperatures are the same. Thus, if enough time is allowed for this transfer of heat to run its course, the temperature a thermometer registers *does* represent the system with which it is in thermal equilibrium. Thermal equilibrium is established when two bodies are in contact with each other and can freely exchange energy.

Furthermore, experimentation has shown that if two systems, A and B, are in thermal equilibrium with each another, and B is in thermal equilibrium with a third system C, then A is also in thermal equilibrium with C. This conclusion may seem obvious, because all three have the same temperature, but it is basic to thermodynamics. It is called the **zeroth law of thermodynamics**.

Note:

The Zeroth Law of Thermodynamics

If two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

This law was postulated in the 1930s, after the first and second laws of thermodynamics had been developed and named. It is called the *zeroth law* because it comes logically before the first and second laws (discussed in [Thermodynamics](#)). Suppose, for example, a cold metal block and a hot metal block are both placed on a metal plate at room temperature. Eventually the cold block and the plate will be in thermal equilibrium. In addition, the hot block and the plate will be in thermal equilibrium. By the zeroth law, we can conclude that the cold block and the hot block are also in thermal equilibrium.

Exercise:
Check Your Understanding

Problem: Does the temperature of a body depend on its size?

Solution:

No, the system can be divided into smaller parts each of which is at the same temperature. We say that the temperature is an *intensive* quantity. Intensive quantities are independent of size.

Section Summary

- Temperature is the quantity measured by a thermometer.
- Temperature is related to the average kinetic energy of atoms and molecules in a system.
- Absolute zero is the temperature at which there is no molecular motion.
- There are three main temperature scales: Celsius, Fahrenheit, and Kelvin.
- Temperatures on one scale can be converted to temperatures on another scale using the following equations:

Equation:

$$T_{\text{°F}} = \frac{9}{5}T_{\text{°C}} + 32$$

Equation:

$$T_{\text{°C}} = \frac{5}{9}(T_{\text{°F}} - 32)$$

Equation:

$$T_{\text{K}} = T_{\text{°C}} + 273.15$$

Equation:

$$T_{\text{°C}} = T_{\text{K}} - 273.15$$

- Systems are in thermal equilibrium when they have the same temperature.
- Thermal equilibrium occurs when two bodies are in contact with each other and can freely exchange energy.
- The zeroth law of thermodynamics states that when two systems, A and B, are in thermal equilibrium with each other, and B is in thermal equilibrium with a third system, C, then A is also in thermal equilibrium with C.

Conceptual Questions

Exercise:

Problem: What does it mean to say that two systems are in thermal equilibrium?

Exercise:

Problem:

Give an example of a physical property that varies with temperature and describe how it is used to measure temperature.

Exercise:

Problem:

When a cold alcohol thermometer is placed in a hot liquid, the column of alcohol goes *down* slightly before going up. Explain why.

Exercise:

Problem:

If you add boiling water to a cup at room temperature, what would you expect the final equilibrium temperature of the unit to be? You will need to include the surroundings as part of the system. Consider the zeroth law of thermodynamics.

Problems & Exercises

Exercise:

Problem: What is the Fahrenheit temperature of a person with a 39.0°C fever?

Solution:

102°F

Exercise:

Problem:

Frost damage to most plants occurs at temperatures of 28.0°F or lower. What is this temperature on the Kelvin scale?

Exercise:

Problem:

To conserve energy, room temperatures are kept at 68.0°F in the winter and 78.0°F in the summer. What are these temperatures on the Celsius scale?

Solution:

20.0°C and 25.6°C

Exercise:**Problem:**

A tungsten light bulb filament may operate at 2900 K . What is its Fahrenheit temperature? What is this on the Celsius scale?

Exercise:**Problem:**

The surface temperature of the Sun is about 5750 K . What is this temperature on the Fahrenheit scale?

Solution:

9890°F

Exercise:**Problem:**

One of the hottest temperatures ever recorded on the surface of Earth was 134°F in Death Valley, CA. What is this temperature in Celsius degrees? What is this temperature in Kelvin?

Exercise:**Problem:**

(a) Suppose a cold front blows into your locale and drops the temperature by 40.0 Fahrenheit degrees. How many degrees Celsius does the temperature decrease when there is a 40.0°F decrease in temperature? (b) Show that any change in temperature in Fahrenheit degrees is nine-fifths the change in Celsius degrees.

Solution:

(a) 22.2°C

$$\begin{aligned}
 \Delta T(^{\circ}\text{F}) &= T_2(^{\circ}\text{F}) - T_1(^{\circ}\text{F}) \\
 \text{(b)} \quad &= \frac{9}{5}T_2(^{\circ}\text{C}) + 32.0^{\circ} - \left(\frac{9}{5}T_1(^{\circ}\text{C}) + 32.0^{\circ}\right) \\
 &= \frac{9}{5}(T_2(^{\circ}\text{C}) - T_1(^{\circ}\text{C})) = \frac{9}{5}\Delta T(^{\circ}\text{C})
 \end{aligned}$$

Exercise:

Problem:

(a) At what temperature do the Fahrenheit and Celsius scales have the same numerical value? (b) At what temperature do the Fahrenheit and Kelvin scales have the same numerical value?

Glossary

temperature

the quantity measured by a thermometer

Celsius scale

temperature scale in which the freezing point of water is 0°C and the boiling point of water is 100°C

degree Celsius

unit on the Celsius temperature scale

Fahrenheit scale

temperature scale in which the freezing point of water is 32°F and the boiling point of water is 212°F

degree Fahrenheit

unit on the Fahrenheit temperature scale

Kelvin scale

temperature scale in which 0 K is the lowest possible temperature, representing absolute zero

absolute zero

the lowest possible temperature; the temperature at which all molecular motion ceases

thermal equilibrium

the condition in which heat no longer flows between two objects that are in contact; the two objects have the same temperature

zeroth law of thermodynamics

law that states that if two objects are in thermal equilibrium, and a third object is in thermal equilibrium with one of those objects, it is also in thermal equilibrium with the other object

Thermal Expansion of Solids and Liquids

- Define and describe thermal expansion.
- Calculate the linear expansion of an object given its initial length, change in temperature, and coefficient of linear expansion.
- Calculate the volume expansion of an object given its initial volume, change in temperature, and coefficient of volume expansion.
- Calculate thermal stress on an object given its original volume, temperature change, volume change, and bulk modulus.



Thermal expansion joints like these in the Auckland Harbour Bridge in New Zealand allow bridges to change length without buckling. (credit: Ingolfson, Wikimedia Commons)

The expansion of alcohol in a thermometer is one of many commonly encountered examples of **thermal expansion**, the change in size or volume of a given mass with temperature. Hot air rises because its volume increases, which causes the hot air's density to be smaller than the density of surrounding air, causing a buoyant (upward) force on the hot air. The same happens in all liquids and gases, driving natural heat transfer upwards in homes, oceans, and weather systems. Solids also undergo thermal expansion. Railroad tracks and bridges, for example, have expansion joints to allow them to freely expand and contract with temperature changes.

What are the basic properties of thermal expansion? First, thermal expansion is clearly related to temperature change. The greater the temperature change, the more a bimetallic strip will bend. Second, it depends on the material. In a thermometer, for example, the expansion of alcohol is much greater than the expansion of the glass containing it.

What is the underlying cause of thermal expansion? As is discussed in [Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature](#), an increase in temperature implies an increase in the kinetic energy of the individual atoms. In a solid, unlike in a gas, the atoms or molecules are closely packed together, but their kinetic energy (in the form of small, rapid vibrations) pushes neighboring atoms or molecules apart from each other. This neighbor-to-neighbor pushing results in a slightly greater distance, on average, between neighbors, and adds up to a larger size for the whole body. For most substances under ordinary conditions, there is no preferred direction, and an increase in temperature will increase the solid's size by a certain fraction in each dimension.

Note:

Linear Thermal Expansion—Thermal Expansion in One Dimension

The change in length ΔL is proportional to length L . The dependence of thermal expansion on temperature, substance, and length is summarized in the equation

Equation:

$$\Delta L = \alpha L \Delta T,$$

where ΔL is the change in length L , ΔT is the change in temperature, and α is the **coefficient of linear expansion**, which varies slightly with temperature.

[\[link\]](#) lists representative values of the coefficient of linear expansion, which may have units of $1/^{\circ}\text{C}$ or $1/\text{K}$. Because the size of a kelvin and a degree Celsius are the same, both α and ΔT can be expressed in units of kelvins or degrees Celsius. The equation $\Delta L = \alpha L \Delta T$ is accurate for small changes in temperature and can be used for large changes in temperature if an average value of α is used.

Material	Coefficient of linear expansion $\alpha(1/^{\circ}\text{C})$	Coefficient of volume expansion $\beta(1/^{\circ}\text{C})$
Solids		
Aluminum	25×10^{-6}	75×10^{-6}
Brass	19×10^{-6}	56×10^{-6}
Copper	17×10^{-6}	51×10^{-6}

Material	Coefficient of linear expansion $\alpha(1/^{\circ}\text{C})$	Coefficient of volume expansion $\beta(1/^{\circ}\text{C})$
Gold	14×10^{-6}	42×10^{-6}
Iron or Steel	12×10^{-6}	35×10^{-6}
Invar (Nickel-iron alloy)	0.9×10^{-6}	2.7×10^{-6}
Lead	29×10^{-6}	87×10^{-6}
Silver	18×10^{-6}	54×10^{-6}
Glass (ordinary)	9×10^{-6}	27×10^{-6}
Glass (Pyrex®)	3×10^{-6}	9×10^{-6}
Quartz	0.4×10^{-6}	1×10^{-6}

Material	Coefficient of linear expansion	Coefficient of volume expansion
	$\alpha(1/^{\circ}\text{C})$	$\beta(1/^{\circ}\text{C})$
Concrete, Brick	$\sim 12 \times 10^{-6}$	$\sim 36 \times 10^{-6}$
Marble (average)	7×10^{-6}	2.1×10^{-5}
Liquids		
Ether		1650×10^{-6}
Ethyl alcohol		1100×10^{-6}
Petrol		950×10^{-6}
Glycerin		500×10^{-6}
Mercury		180×10^{-6}

Material	Coefficient of linear expansion $\alpha(1/^{\circ}\text{C})$	Coefficient of volume expansion $\beta(1/^{\circ}\text{C})$
Water		210×10^{-6}
Gases		
Air and most other gases at atmospheric pressure		3400×10^{-6}

Thermal Expansion Coefficients at 20°C[\[footnote\]](#)

Values for liquids and gases are approximate.

Example:

Calculating Linear Thermal Expansion: The Golden Gate Bridge

The main span of San Francisco's Golden Gate Bridge is 1275 m long at its coldest. The bridge is exposed to temperatures ranging from -15°C to 40°C . What is its change in length between these temperatures? Assume that the bridge is made entirely of steel.

Strategy

Use the equation for linear thermal expansion $\Delta L = \alpha L \Delta T$ to calculate the change in length, ΔL . Use the coefficient of linear expansion, α , for steel from [\[link\]](#), and note that the change in temperature, ΔT , is 55°C .

Solution

Plug all of the known values into the equation to solve for ΔL .

Equation:

$$\Delta L = \alpha L \Delta T = \left(\frac{12 \times 10^{-6}}{^{\circ}\text{C}} \right) (1275 \text{ m}) (55^{\circ}\text{C}) = 0.84 \text{ m}.$$

Discussion

Although not large compared with the length of the bridge, this change in length is observable. It is generally spread over many expansion joints so that the expansion at each joint is small.

Thermal Expansion in Two and Three Dimensions

Objects expand in all dimensions, as illustrated in [\[link\]](#). That is, their areas and volumes, as well as their lengths, increase with temperature. Holes also get larger with temperature. If you cut a hole in a metal plate, the remaining material will expand exactly as it would if the plug was still in place. The plug would get bigger, and so the hole must get bigger too. (Think of the ring of neighboring atoms or molecules on the wall of the hole as pushing each other farther apart as temperature increases. Obviously, the ring of neighbors must get slightly larger, so the hole gets slightly larger).

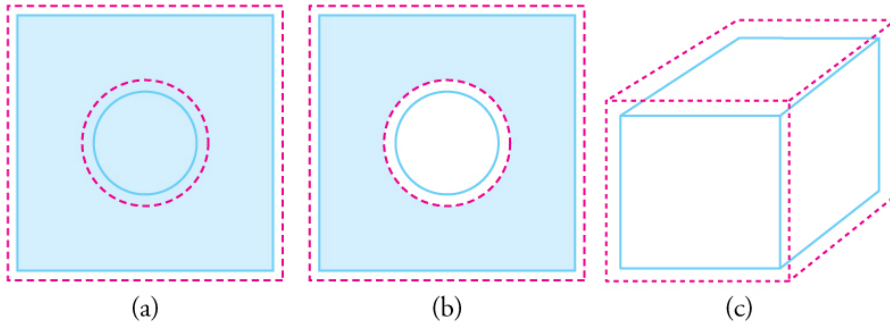
Note:**Thermal Expansion in Two Dimensions**

For small temperature changes, the change in area ΔA is given by

Equation:

$$\Delta A = 2\alpha A\Delta T,$$

where ΔA is the change in area A , ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.



In general, objects expand in all directions as temperature increases. In these drawings, the original boundaries of the objects are shown with solid lines, and the expanded boundaries with dashed lines. (a) Area increases because both length and width increase. The area of a circular plug also increases. (b) If the plug is removed, the hole it leaves becomes larger with increasing temperature, just as if the expanding plug were still in place. (c) Volume also increases, because all three dimensions increase.

Note:

Thermal Expansion in Three Dimensions

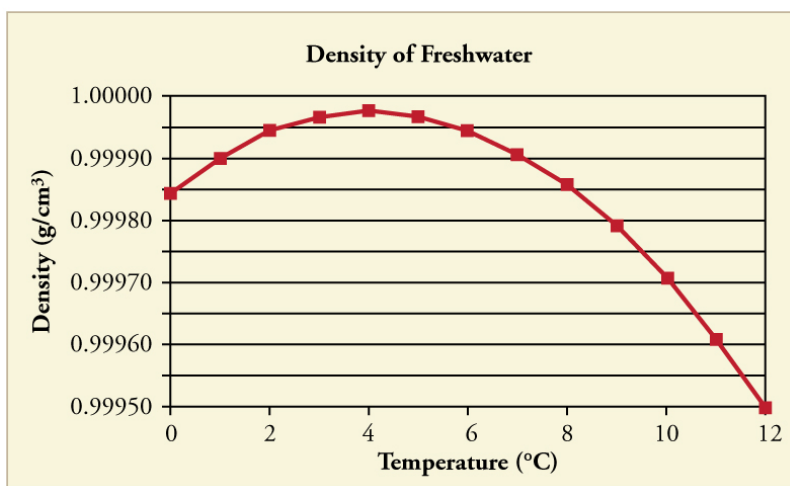
The change in volume ΔV is very nearly $\Delta V = 3\alpha V \Delta T$. This equation is usually written as

Equation:

$$\Delta V = \beta V \Delta T,$$

where β is the **coefficient of volume expansion** and $\beta \approx 3\alpha$. Note that the values of β in [\[link\]](#) are almost exactly equal to 3α .

In general, objects will expand with increasing temperature. Water is the most important exception to this rule. Water expands with increasing temperature (its density *decreases*) when it is at temperatures greater than 4°C (40°F). However, it expands with *decreasing* temperature when it is between $+4^{\circ}\text{C}$ and 0°C (40°F to 32°F). Water is densest at $+4^{\circ}\text{C}$. (See [\[link\]](#).) Perhaps the most striking effect of this phenomenon is the freezing of water in a pond. When water near the surface cools down to 4°C it is denser than the remaining water and thus will sink to the bottom. This “turnover” results in a layer of warmer water near the surface, which is then cooled. Eventually the pond has a uniform temperature of 4°C . If the temperature in the surface layer drops below 4°C , the water is less dense than the water below, and thus stays near the top. As a result, the pond surface can completely freeze over. The ice on top of liquid water provides an insulating layer from winter’s harsh exterior air temperatures. Fish and other aquatic life can survive in 4°C water beneath ice, due to this unusual characteristic of water. It also produces circulation of water in the pond that is necessary for a healthy ecosystem of the body of water.



The density of water as a function of temperature. Note that the thermal expansion is actually very small. The maximum density at $+4^{\circ}\text{C}$ is only 0.0075% greater than the density at 2°C , and 0.012% greater than that at 0°C .

Note:**Making Connections: Real-World Connections—Filling the Tank**

Differences in the thermal expansion of materials can lead to interesting effects at the gas station. One example is the dripping of gasoline from a freshly filled tank on a hot day. Gasoline starts out at the temperature of the ground under the gas station, which is cooler than the air temperature above. The gasoline cools the steel tank when it is filled. Both gasoline and steel tank expand as they warm to air temperature, but gasoline expands much more than steel, and so it may overflow.

This difference in expansion can also cause problems when interpreting the gasoline gauge. The actual amount (mass) of gasoline left in the tank when the gauge hits “empty” is a lot less in the summer than in the winter. The gasoline has the same volume as it does in the winter when the “add fuel” light goes on, but because the gasoline has expanded, there is less mass. If you are used to getting another 40 miles on “empty” in the winter, beware—you will probably run out much more quickly in the summer.



Because the gas expands more than the gas tank with increasing temperature, you can't drive as many miles on “empty” in the summer as you can in the winter.

(credit: Hector Alejandro,
Flickr)

Example:

Calculating Thermal Expansion: Gas vs. Gas Tank

Suppose your 60.0-L (15.9-gal) steel gasoline tank is full of gas, so both the tank and the gasoline have a temperature of 15.0°C. How much gasoline has spilled by the time they warm to 35.0°C?

Strategy

The tank and gasoline increase in volume, but the gasoline increases more, so the amount spilled is the difference in their volume changes. (The gasoline tank can be treated as solid steel.) We can use the equation for volume expansion to calculate the change in volume of the gasoline and of the tank.

Solution

1. Use the equation for volume expansion to calculate the increase in volume of the steel tank:

Equation:

$$\Delta V_s = \beta_s V_s \Delta T.$$

2. The increase in volume of the gasoline is given by this equation:

Equation:

$$\Delta V_{\text{gas}} = \beta_{\text{gas}} V_{\text{gas}} \Delta T.$$

3. Find the difference in volume to determine the amount spilled as

Equation:

$$V_{\text{spill}} = \Delta V_{\text{gas}} - \Delta V_s.$$

Alternatively, we can combine these three equations into a single equation. (Note that the original volumes are equal.)

Equation:

$$\begin{aligned}
 V_{\text{spill}} &= (\beta_{\text{gas}} - \beta_{\text{s}})V\Delta T \\
 &= [(950 - 35) \times 10^{-6} / ^\circ\text{C}] (60.0 \text{ L})(20.0^\circ\text{C}) \\
 &= 1.10 \text{ L.}
 \end{aligned}$$

Discussion

This amount is significant, particularly for a 60.0-L tank. The effect is so striking because the gasoline and steel expand quickly. The rate of change in thermal properties is discussed in [Heat and Heat Transfer Methods](#).

If you try to cap the tank tightly to prevent overflow, you will find that it leaks anyway, either around the cap or by bursting the tank. Tightly constricting the expanding gas is equivalent to compressing it, and both liquids and solids resist being compressed with extremely large forces. To avoid rupturing rigid containers, these containers have air gaps, which allow them to expand and contract without stressing them.

Thermal Stress

Thermal stress is created by thermal expansion or contraction (see [Elasticity: Stress and Strain](#) for a discussion of stress and strain). Thermal stress can be destructive, such as when expanding gasoline ruptures a tank. It can also be useful, for example, when two parts are joined together by heating one in manufacturing, then slipping it over the other and allowing the combination to cool. Thermal stress can explain many phenomena, such as the weathering of rocks and pavement by the expansion of ice when it freezes.

Example:

Calculating Thermal Stress: Gas Pressure

What pressure would be created in the gasoline tank considered in [\[link\]](#), if the gasoline increases in temperature from 15.0°C to 35.0°C without being allowed to expand? Assume that the bulk modulus B for gasoline is $1.00 \times 10^9 \text{ N/m}^2$. (For more on bulk modulus, see [Elasticity: Stress and Strain](#).)

Strategy

To solve this problem, we must use the following equation, which relates a change in volume ΔV to pressure:

Equation:

$$\Delta V = \frac{1}{B} \frac{F}{A} V_0,$$

where F/A is pressure, V_0 is the original volume, and B is the bulk modulus of the material involved. We will use the amount spilled in [\[link\]](#) as the change in volume, ΔV .

Solution

1. Rearrange the equation for calculating pressure:

Equation:

$$P = \frac{F}{A} = \frac{\Delta V}{V_0} B.$$

2. Insert the known values. The bulk modulus for gasoline is $B = 1.00 \times 10^9 \text{ N/m}^2$. In the previous example, the change in volume $\Delta V = 1.10 \text{ L}$ is the amount that would spill. Here, $V_0 = 60.0 \text{ L}$ is the original volume of the gasoline. Substituting these values into the equation, we obtain

Equation:

$$P = \frac{1.10 \text{ L}}{60.0 \text{ L}} (1.00 \times 10^9 \text{ Pa}) = 1.83 \times 10^7 \text{ Pa}.$$

Discussion

This pressure is about 2500 lb/in^2 , *much* more than a gasoline tank can handle.

Forces and pressures created by thermal stress are typically as great as that in the example above. Railroad tracks and roadways can buckle on hot days if they lack sufficient expansion joints. (See [\[link\]](#).) Power lines sag more in the summer than in the winter, and will snap in cold weather if there is

insufficient slack. Cracks open and close in plaster walls as a house warms and cools. Glass cooking pans will crack if cooled rapidly or unevenly, because of differential contraction and the stresses it creates. (Pyrex® is less susceptible because of its small coefficient of thermal expansion.) Nuclear reactor pressure vessels are threatened by overly rapid cooling, and although none have failed, several have been cooled faster than considered desirable. Biological cells are ruptured when foods are frozen, detracting from their taste. Repeated thawing and freezing accentuate the damage. Even the oceans can be affected. A significant portion of the rise in sea level that is resulting from global warming is due to the thermal expansion of sea water.



Thermal stress contributes to the formation of potholes.
(credit: Editor5807, Wikimedia Commons)

Metal is regularly used in the human body for hip and knee implants. Most implants need to be replaced over time because, among other things, metal does not bond with bone. Researchers are trying to find better metal coatings that would allow metal-to-bone bonding. One challenge is to find a coating that has an expansion coefficient similar to that of metal. If the

expansion coefficients are too different, the thermal stresses during the manufacturing process lead to cracks at the coating-metal interface.

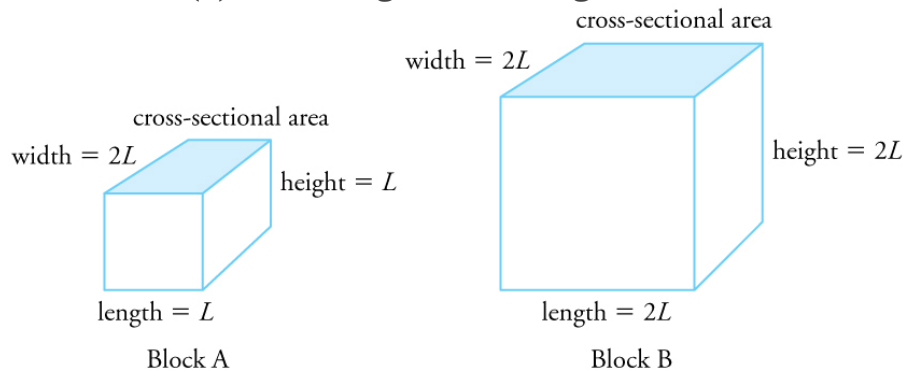
Another example of thermal stress is found in the mouth. Dental fillings can expand differently from tooth enamel. It can give pain when eating ice cream or having a hot drink. Cracks might occur in the filling. Metal fillings (gold, silver, etc.) are being replaced by composite fillings (porcelain), which have smaller coefficients of expansion, and are closer to those of teeth.

Exercise:

Check Your Understanding

Problem:

Two blocks, A and B, are made of the same material. Block A has dimensions $l \times w \times h = L \times 2L \times L$ and Block B has dimensions $2L \times 2L \times 2L$. If the temperature changes, what is (a) the change in the volume of the two blocks, (b) the change in the cross-sectional area $l \times w$, and (c) the change in the height h of the two blocks?



Solution:

(a) The change in volume is proportional to the original volume. Block A has a volume of $L \times 2L \times L = 2L^3$. Block B has a volume of $2L \times 2L \times 2L = 8L^3$, which is 4 times that of Block A. Thus the change in volume of Block B should be 4 times the change in volume of Block A.

(b) The change in area is proportional to the area. The cross-sectional area of Block A is $L \times 2L = 2L^2$, while that of Block B is

$2L \times 2L = 4L^2$. Because cross-sectional area of Block B is twice that of Block A, the change in the cross-sectional area of Block B is twice that of Block A.

(c) The change in height is proportional to the original height. Because the original height of Block B is twice that of A, the change in the height of Block B is twice that of Block A.

Section Summary

- Thermal expansion is the increase, or decrease, of the size (length, area, or volume) of a body due to a change in temperature.
- Thermal expansion is large for gases, and relatively small, but not negligible, for liquids and solids.
- Linear thermal expansion is

Equation:

$$\Delta L = \alpha L \Delta T,$$

where ΔL is the change in length L , ΔT is the change in temperature, and α is the coefficient of linear expansion, which varies slightly with temperature.

- The change in area due to thermal expansion is

Equation:

$$\Delta A = 2\alpha A \Delta T,$$

where ΔA is the change in area.

- The change in volume due to thermal expansion is

Equation:

$$\Delta V = \beta V \Delta T,$$

where β is the coefficient of volume expansion and $\beta \approx 3\alpha$. Thermal stress is created when thermal expansion is constrained.

Conceptual Questions

Exercise:**Problem:**

Thermal stresses caused by uneven cooling can easily break glass cookware. Explain why Pyrex®, a glass with a small coefficient of linear expansion, is less susceptible.

Exercise:**Problem:**

Water expands significantly when it freezes: a volume increase of about 9% occurs. As a result of this expansion and because of the formation and growth of crystals as water freezes, anywhere from 10% to 30% of biological cells are burst when animal or plant material is frozen. Discuss the implications of this cell damage for the prospect of preserving human bodies by freezing so that they can be thawed at some future date when it is hoped that all diseases are curable.

Exercise:**Problem:**

One method of getting a tight fit, say of a metal peg in a hole in a metal block, is to manufacture the peg slightly larger than the hole. The peg is then inserted when at a different temperature than the block. Should the block be hotter or colder than the peg during insertion? Explain your answer.

Exercise:**Problem:**

Does it really help to run hot water over a tight metal lid on a glass jar before trying to open it? Explain your answer.

Exercise:

Problem:

Liquids and solids expand with increasing temperature, because the kinetic energy of a body's atoms and molecules increases. Explain why some materials *shrink* with increasing temperature.

Problems & Exercises**Exercise:****Problem:**

The height of the Washington Monument is measured to be 170 m on a day when the temperature is 35.0°C . What will its height be on a day when the temperature falls to -10.0°C ? Although the monument is made of limestone, assume that its thermal coefficient of expansion is the same as marble's.

Solution:

169.98 m

Exercise:**Problem:**

How much taller does the Eiffel Tower become at the end of a day when the temperature has increased by 15°C ? Its original height is 321 m and you can assume it is made of steel.

Exercise:**Problem:**

What is the change in length of a 3.00-cm-long column of mercury if its temperature changes from 37.0°C to 40.0°C , assuming the mercury is unconstrained?

Solution:

$$5.4 \times 10^{-6} \text{ m}$$

Exercise:**Problem:**

How large an expansion gap should be left between steel railroad rails if they may reach a maximum temperature 35.0°C greater than when they were laid? Their original length is 10.0 m.

Exercise:**Problem:**

You are looking to purchase a small piece of land in Hong Kong. The price is “only” \$60,000 per square meter! The land title says the dimensions are $20 \text{ m} \times 30 \text{ m}$. By how much would the total price change if you measured the parcel with a steel tape measure on a day when the temperature was 20°C above normal?

Solution:

Because the area gets smaller, the price of the land DECREASES by ~\$17,000.

Exercise:**Problem:**

Global warming will produce rising sea levels partly due to melting ice caps but also due to the expansion of water as average ocean temperatures rise. To get some idea of the size of this effect, calculate the change in length of a column of water 1.00 km high for a temperature increase of 1.00°C . Note that this calculation is only approximate because ocean warming is not uniform with depth.

Exercise:**Problem:**

Show that 60.0 L of gasoline originally at 15.0°C will expand to 61.1 L when it warms to 35.0°C , as claimed in [\[link\]](#).

Solution:

Equation:

$$\begin{aligned} V &= V_0 + \Delta V = V_0(1 + \beta\Delta T) \\ &= (60.00 \text{ L})[1 + (950 \times 10^{-6}/^{\circ}\text{C})(35.0^{\circ}\text{C} - 15.0^{\circ}\text{C})] \\ &= 61.1 \text{ L} \end{aligned}$$

Exercise:

Problem:

(a) Suppose a meter stick made of steel and one made of invar (an alloy of iron and nickel) are the same length at 0°C . What is their difference in length at 22.0°C ? (b) Repeat the calculation for two 30.0-m-long surveyor's tapes.

Exercise:

Problem:

(a) If a 500-mL glass beaker is filled to the brim with ethyl alcohol at a temperature of 5.00°C , how much will overflow when its temperature reaches 22.0°C ? (b) How much less water would overflow under the same conditions?

Solution:

(a) 9.35 mL

(b) 7.56 mL

Exercise:

Problem:

Most automobiles have a coolant reservoir to catch radiator fluid that may overflow when the engine is hot. A radiator is made of copper and is filled to its 16.0-L capacity when at 10.0°C. What volume of radiator fluid will overflow when the radiator and fluid reach their 95.0°C operating temperature, given that the fluid's volume coefficient of expansion is $\beta = 400 \times 10^{-6} / ^\circ\text{C}$? Note that this coefficient is approximate, because most car radiators have operating temperatures of greater than 95.0°C.

Exercise:**Problem:**

A physicist makes a cup of instant coffee and notices that, as the coffee cools, its level drops 3.00 mm in the glass cup. Show that this decrease cannot be due to thermal contraction by calculating the decrease in level if the 350 cm³ of coffee is in a 7.00-cm-diameter cup and decreases in temperature from 95.0°C to 45.0°C. (Most of the drop in level is actually due to escaping bubbles of air.)

Solution:

0.832 mm

Exercise:**Problem:**

(a) The density of water at 0°C is very nearly 1000 kg/m³ (it is actually 999.84 kg/m³), whereas the density of ice at 0°C is 917 kg/m³. Calculate the pressure necessary to keep ice from expanding when it freezes, neglecting the effect such a large pressure would have on the freezing temperature. (This problem gives you only an indication of how large the forces associated with freezing water might be.) (b) What are the implications of this result for biological cells that are frozen?

Exercise:**Problem:**

Show that $\beta \approx 3\alpha$, by calculating the change in volume ΔV of a cube with sides of length L .

Solution:

We know how the length changes with temperature: $\Delta L = \alpha L_0 \Delta T$. Also we know that the volume of a cube is related to its length by $V = L^3$, so the final volume is then $V = V_0 + \Delta V = (L_0 + \Delta L)^3$. Substituting for ΔL gives

Equation:

$$V = (L_0 + \alpha L_0 \Delta T)^3 = L_0^3 (1 + \alpha \Delta T)^3.$$

Now, because $\alpha \Delta T$ is small, we can use the binomial expansion:

Equation:

$$V \approx L_0^3 (1 + 3\alpha \Delta T) = L_0^3 + 3\alpha L_0^3 \Delta T.$$

So writing the length terms in terms of volumes gives $V = V_0 + \Delta V \approx V_0 + 3\alpha V_0 \Delta T$, and so

Equation:

$$\Delta V = \beta V_0 \Delta T \approx 3\alpha V_0 \Delta T, \text{ or } \beta \approx 3\alpha.$$

Glossary

thermal expansion

the change in size or volume of an object with change in temperature

coefficient of linear expansion

α , the change in length, per unit length, per 1°C change in temperature; a constant used in the calculation of linear expansion; the coefficient of linear expansion depends on the material and to some degree on the temperature of the material

coefficient of volume expansion

β , the change in volume, per unit volume, per 1°C change in temperature

thermal stress

stress caused by thermal expansion or contraction

The Ideal Gas Law

- State the ideal gas law in terms of molecules and in terms of moles.
- Use the ideal gas law to calculate pressure change, temperature change, volume change, or the number of molecules or moles in a given volume.
- Use Avogadro's number to convert between number of molecules and number of moles.

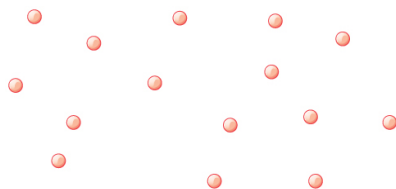


The air inside this hot air balloon flying over Putrajaya, Malaysia, is hotter than the ambient air. As a result, the balloon experiences a buoyant force pushing it upward.
(credit: Kevin Poh, Flickr)

In this section, we continue to explore the thermal behavior of gases. In particular, we examine the characteristics of atoms and molecules that compose gases. (Most gases, for example nitrogen, N_2 , and oxygen, O_2 , are composed of two or more atoms. We will primarily use the term “molecule” in discussing a gas because the term can also be applied to monatomic gases, such as helium.)

Gases are easily compressed. We can see evidence of this in [\[link\]](#), where you will note that gases have the *largest* coefficients of volume expansion. The large coefficients mean that gases expand and contract very rapidly with temperature changes. In addition, you will note that most gases expand at the *same* rate, or have the same β . This raises the question as to why gases should all act in nearly the same way, when liquids and solids have widely varying expansion rates.

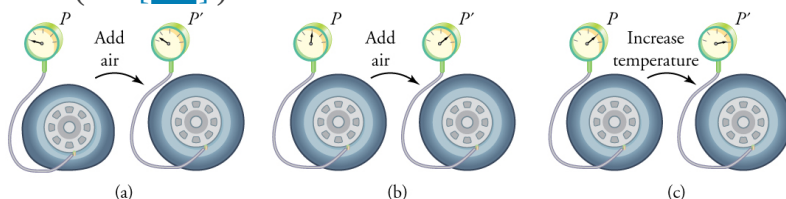
The answer lies in the large separation of atoms and molecules in gases, compared to their sizes, as illustrated in [\[link\]](#). Because atoms and molecules have large separations, forces between them can be ignored, except when they collide with each other during collisions. The motion of atoms and molecules (at temperatures well above the boiling temperature) is fast, such that the gas occupies all of the accessible volume and the expansion of gases is rapid. In contrast, in liquids and solids, atoms and molecules are closer together and are quite sensitive to the forces between them.



Atoms and molecules in a gas are typically widely separated, as shown.

Because the forces between them are quite weak at these distances, the properties of a gas depend more on the number of atoms per unit volume and on temperature than on the type of atom.

To get some idea of how pressure, temperature, and volume of a gas are related to one another, consider what happens when you pump air into an initially deflated tire. The tire's volume first increases in direct proportion to the amount of air injected, without much increase in the tire pressure. Once the tire has expanded to nearly its full size, the walls limit volume expansion. If we continue to pump air into it, the pressure increases. The pressure will further increase when the car is driven and the tires move. Most manufacturers specify optimal tire pressure for cold tires. (See [\[link\]](#).)



(a) When air is pumped into a deflated tire, its volume first increases without much increase in pressure. (b) When the tire is filled to a certain point, the tire walls resist further expansion and the pressure increases with

more air. (c) Once the tire is inflated, its pressure increases with temperature.

At room temperatures, collisions between atoms and molecules can be ignored. In this case, the gas is called an ideal gas, in which case the relationship between the pressure, volume, and temperature is given by the equation of state called the ideal gas law.

Note:

Ideal Gas Law

The **ideal gas law** states that

Equation:

$$PV = NkT,$$

where P is the absolute pressure of a gas, V is the volume it occupies, N is the number of atoms and molecules in the gas, and T is its absolute temperature. The constant k is called the **Boltzmann constant** in honor of Austrian physicist Ludwig Boltzmann (1844–1906) and has the value

Equation:

$$k = 1.38 \times 10^{-23} \text{ J/K}.$$

The ideal gas law can be derived from basic principles, but was originally deduced from experimental measurements of Charles' law (that volume occupied by a gas is proportional to temperature at a fixed pressure) and from Boyle's law (that for a fixed temperature, the product PV is a constant). In the ideal gas model, the volume occupied by its atoms and molecules is a negligible fraction of V . The ideal gas law describes the behavior of real gases under most conditions. (Note, for example, that N is the total number of atoms and molecules, independent of the type of gas.)

Let us see how the ideal gas law is consistent with the behavior of filling the tire when it is pumped slowly and the temperature is constant. At first, the pressure P is essentially equal to atmospheric pressure, and the volume V increases in direct proportion to the number of atoms and molecules N put into the tire. Once the volume of the tire is constant, the equation $PV = NkT$ predicts that the pressure should increase in proportion to *the number N of atoms and molecules*.

Example:

Calculating Pressure Changes Due to Temperature Changes: Tire Pressure

Suppose your bicycle tire is fully inflated, with an absolute pressure of 7.00×10^5 Pa (a gauge pressure of just under 90.0 lb/in²) at a temperature of 18.0°C. What is the pressure after its temperature has risen to 35.0°C? Assume that there are no appreciable leaks or changes in volume.

Strategy

The pressure in the tire is changing only because of changes in temperature. First we need to identify what we know and what we want to know, and then identify an equation to solve for the unknown.

We know the initial pressure $P_0 = 7.00 \times 10^5$ Pa, the initial temperature $T_0 = 18.0^\circ\text{C}$, and the final temperature $T_f = 35.0^\circ\text{C}$. We must find the final pressure P_f . How can we use the equation $PV = NkT$? At first, it may seem that not enough information is given, because the volume V and number of atoms N are not specified. What we can do is use the equation twice: $P_0V_0 = NkT_0$ and $P_fV_f = NkT_f$. If we divide P_fV_f by P_0V_0 we can come up with an equation that allows us to solve for P_f .

Equation:

$$\frac{P_f V_f}{P_0 V_0} = \frac{N_f k T_f}{N_0 k T_0}$$

Since the volume is constant, V_f and V_0 are the same and they cancel out. The same is true for N_f and N_0 , and k , which is a constant. Therefore,

Equation:

$$\frac{P_f}{P_0} = \frac{T_f}{T_0}.$$

We can then rearrange this to solve for P_f :

Equation:

$$P_f = P_0 \frac{T_f}{T_0},$$

where the temperature must be in units of kelvins, because T_0 and T_f are absolute temperatures.

Solution

1. Convert temperatures from Celsius to Kelvin.

Equation:

$$\begin{aligned} T_0 &= (18.0 + 273)\text{K} = 291 \text{ K} \\ T_f &= (35.0 + 273)\text{K} = 308 \text{ K} \end{aligned}$$

2. Substitute the known values into the equation.

Equation:

$$P_f = P_0 \frac{T_f}{T_0} = 7.00 \times 10^5 \text{ Pa} \left(\frac{308 \text{ K}}{291 \text{ K}} \right) = 7.41 \times 10^5 \text{ Pa}$$

Discussion

The final temperature is about 6% greater than the original temperature, so the final pressure is about 6% greater as well. Note that *absolute* pressure and *absolute* temperature must be used in the ideal gas law.

Note:

Making Connections: Take-Home Experiment—Refrigerating a Balloon

Inflate a balloon at room temperature. Leave the inflated balloon in the refrigerator overnight. What happens to the balloon, and why?

Example:

Calculating the Number of Molecules in a Cubic Meter of Gas

How many molecules are in a typical object, such as gas in a tire or water in a drink? We can use the ideal gas law to give us an idea of how large N typically is.

Calculate the number of molecules in a cubic meter of gas at standard temperature and pressure (STP), which is defined to be 0°C and atmospheric pressure.

Strategy

Because pressure, volume, and temperature are all specified, we can use the ideal gas law $PV = NkT$, to find N .

Solution

1. Identify the knowns.

Equation:

$$T = 0^\circ\text{C} = 273 \text{ K}$$

$$P = 1.01 \times 10^5 \text{ Pa}$$

$$V = 1.00 \text{ m}^3$$

$$k = 1.38 \times 10^{-23} \text{ J/K}$$

2. Identify the unknown: number of molecules, N .

3. Rearrange the ideal gas law to solve for N .

Equation:

$$PV = NkT$$

$$N = \frac{PV}{kT}$$

4. Substitute the known values into the equation and solve for N .

Equation:

$$N = \frac{PV}{kT} = \frac{(1.01 \times 10^5 \text{ Pa})(1.00 \text{ m}^3)}{(1.38 \times 10^{-23} \text{ J/K})(273 \text{ K})} = 2.68 \times 10^{25} \text{ molecules}$$

Discussion

This number is undeniably large, considering that a gas is mostly empty space. N is huge, even in small volumes. For example, 1 cm^3 of a gas at STP has 2.68×10^{19} molecules in it. Once again, note that N is the same for all types or mixtures of gases.

Moles and Avogadro's Number

It is sometimes convenient to work with a unit other than molecules when measuring the amount of substance. A **mole** (abbreviated mol) is defined to be the amount of a substance that contains as many atoms or molecules as there are atoms in exactly 12 grams (0.012 kg) of carbon-12. The actual number of atoms or molecules in one mole is called **Avogadro's number** (N_A), in recognition of Italian scientist Amedeo Avogadro (1776–1856). He developed the concept of the mole, based on the hypothesis that equal volumes of gas, at the same pressure and temperature, contain equal numbers of molecules. That is, the number is independent of the type of gas. This hypothesis has been confirmed, and the value of Avogadro's number is

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}.$$

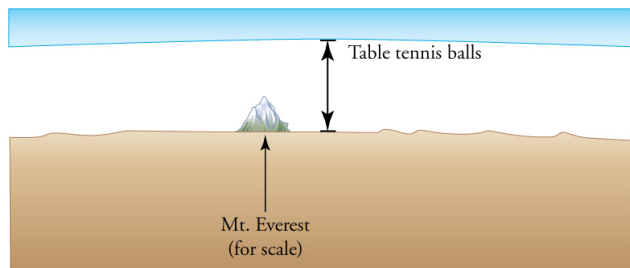
Note:

Avogadro's Number

One mole always contains 6.02×10^{23} particles (atoms or molecules), independent of the element or substance. A mole of any substance has a mass in grams equal to its molecular mass, which can be calculated from the atomic masses given in the periodic table of elements.

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}$$



How big is a mole? On a macroscopic level, one mole of table tennis balls would cover the Earth to a depth of about 40 km.

Exercise:
Check Your Understanding

Problem:

The active ingredient in a Tylenol pill is 325 mg of acetaminophen ($\text{C}_8\text{H}_9\text{NO}_2$). Find the number of active molecules of acetaminophen in a single pill.

Solution:

We first need to calculate the molar mass (the mass of one mole) of acetaminophen. To do this, we need to multiply the number of atoms of each element by the element's atomic mass.

Equation:

$$(8 \text{ moles of carbon})(12 \text{ grams/mole}) + (9 \text{ moles hydrogen})(1 \text{ gram/mole}) \\ + (1 \text{ mole nitrogen})(14 \text{ grams/mole}) + (2 \text{ moles oxygen})(16 \text{ grams/mole}) = 151 \text{ g}$$

Then we need to calculate the number of moles in 325 mg.

Equation:

$$\left(\frac{325 \text{ mg}}{151 \text{ grams/mole}} \right) \left(\frac{1 \text{ gram}}{1000 \text{ mg}} \right) = 2.15 \times 10^{-3} \text{ moles}$$

Then use Avogadro's number to calculate the number of molecules.

Equation:

$$N = (2.15 \times 10^{-3} \text{ moles}) (6.02 \times 10^{23} \text{ molecules/mole}) = 1.30 \times 10^{21} \text{ molecules}$$

Example:

Calculating Moles per Cubic Meter and Liters per Mole

Calculate: (a) the number of moles in 1.00 m^3 of gas at STP, and (b) the number of liters of gas per mole.

Strategy and Solution

(a) We are asked to find the number of moles per cubic meter, and we know from [\[link\]](#) that the number of molecules per cubic meter at STP is 2.68×10^{25} . The number of moles can be found by dividing the number of molecules by Avogadro's number. We let n stand for the number of moles,

Equation:

$$n \text{ mol/m}^3 = \frac{N \text{ molecules/m}^3}{6.02 \times 10^{23} \text{ molecules/mol}} = \frac{2.68 \times 10^{25} \text{ molecules/m}^3}{6.02 \times 10^{23} \text{ molecules/mol}} = 44.5 \text{ mol/m}^3.$$

(b) Using the value obtained for the number of moles in a cubic meter, and converting cubic meters to liters, we obtain

Equation:

$$\frac{(10^3 \text{ L/m}^3)}{44.5 \text{ mol/m}^3} = 22.5 \text{ L/mol.}$$

Discussion

This value is very close to the accepted value of 22.4 L/mol. The slight difference is due to rounding errors caused by using three-digit input. Again this number is the same for all gases. In other words, it is independent of the gas.

The (average) molar weight of air (approximately 80% N₂ and 20% O₂) is $M = 28.8 \text{ g}$. Thus the mass of one cubic meter of air is 1.28 kg. If a living room has dimensions $5 \text{ m} \times 5 \text{ m} \times 3 \text{ m}$, the mass of air inside the room is 96 kg, which is the typical mass of a human.

Exercise:

Check Your Understanding

Problem:

The density of air at standard conditions ($P = 1 \text{ atm}$ and $T = 20^\circ\text{C}$) is 1.28 kg/m^3 . At what pressure is the density 0.64 kg/m^3 if the temperature and number of molecules are kept constant?

Solution:

The best way to approach this question is to think about what is happening. If the density drops to half its original value and no molecules are lost, then the volume must double. If we look at the equation $PV = NkT$, we see that when the temperature is constant, the pressure is inversely proportional to volume. Therefore, if the volume doubles, the pressure must drop to half its original value, and $P_f = 0.50 \text{ atm}$.

The Ideal Gas Law Restated Using Moles

A very common expression of the ideal gas law uses the number of moles, n , rather than the number of atoms and molecules, N . We start from the ideal gas law,

Equation:

$$PV = NkT,$$

and multiply and divide the equation by Avogadro's number N_A . This gives

Equation:

$$PV = \frac{N}{N_A} N_A k T.$$

Note that $n = N/N_A$ is the number of moles. We define the universal gas constant $R = N_A k$, and obtain the ideal gas law in terms of moles.

Note:

Ideal Gas Law (in terms of moles)

The ideal gas law (in terms of moles) is

Equation:

$$PV = nRT.$$

The numerical value of R in SI units is

Equation:

$$R = N_A k = (6.02 \times 10^{23} \text{ mol}^{-1})(1.38 \times 10^{-23} \text{ J/K}) = 8.31 \text{ J/mol} \cdot \text{K}.$$

In other units,

Equation:

$$R = 1.99 \text{ cal/mol} \cdot \text{K}$$

$$R = 0.0821 \text{ L} \cdot \text{atm/mol} \cdot \text{K}.$$

You can use whichever value of R is most convenient for a particular problem.

Example:

Calculating Number of Moles: Gas in a Bike Tire

How many moles of gas are in a bike tire with a volume of $2.00 \times 10^{-3} \text{ m}^3$ (2.00 L), a pressure of $7.00 \times 10^5 \text{ Pa}$ (a gauge pressure of just under 90.0 lb/in²), and at a temperature of 18.0°C?

Strategy

Identify the knowns and unknowns, and choose an equation to solve for the unknown. In this case, we solve the ideal gas law, $PV = nRT$, for the number of moles n .

Solution

1. Identify the knowns.

Equation:

$$P = 7.00 \times 10^5 \text{ Pa}$$

$$V = 2.00 \times 10^{-3} \text{ m}^3$$

$$T = 18.0^\circ\text{C} = 291 \text{ K}$$

$$R = 8.31 \text{ J/mol} \cdot \text{K}$$

2. Rearrange the equation to solve for n and substitute known values.

Equation:

$$\begin{aligned}n &= \frac{PV}{RT} = \frac{(7.00 \times 10^5 \text{ Pa})(2.00 \times 10^{-3} \text{ m}^3)}{(8.31 \text{ J/mol}\cdot\text{K})(291 \text{ K})} \\&= 0.579 \text{ mol}\end{aligned}$$

Discussion

The most convenient choice for R in this case is $8.31 \text{ J/mol}\cdot\text{K}$, because our known quantities are in SI units. The pressure and temperature are obtained from the initial conditions in [\[link\]](#), but we would get the same answer if we used the final values.

The ideal gas law can be considered to be another manifestation of the law of conservation of energy (see [Conservation of Energy](#)). Work done on a gas results in an increase in its energy, increasing pressure and/or temperature, or decreasing volume. This increased energy can also be viewed as increased internal kinetic energy, given the gas's atoms and molecules.

The Ideal Gas Law and Energy

Let us now examine the role of energy in the behavior of gases. When you inflate a bike tire by hand, you do work by repeatedly exerting a force through a distance. This energy goes into increasing the pressure of air inside the tire and increasing the temperature of the pump and the air.

The ideal gas law is closely related to energy: the units on both sides are joules. The right-hand side of the ideal gas law in $PV = NkT$ is NkT . This term is roughly the amount of translational kinetic energy of N atoms or molecules at an absolute temperature T , as we shall see formally in [Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature](#). The left-hand side of the ideal gas law is PV , which also has the units of joules. We know from our study of fluids that pressure is one type of potential energy per unit volume, so pressure multiplied by volume is energy. The important point is that there is energy in a gas related to both its pressure and its volume. The energy can be changed when the gas is doing work as it expands—something we explore in [Heat and Heat Transfer Methods](#)—similar to what occurs in gasoline or steam engines and turbines.

Note:

Problem-Solving Strategy: The Ideal Gas Law

Step 1 Examine the situation to determine that an ideal gas is involved. Most gases are nearly ideal.

Step 2 Make a list of what quantities are given, or can be inferred from the problem as stated (identify the known quantities). Convert known values into proper SI units (K for temperature, Pa for pressure, m^3 for volume, molecules for N , and moles for n).

Step 3 Identify exactly what needs to be determined in the problem (identify the unknown quantities). A written list is useful.

Step 4 Determine whether the number of molecules or the number of moles is known, in order to decide which form of the ideal gas law to use. The first form is $PV = NkT$ and involves N , the number of atoms or molecules. The second form is $PV = nRT$ and involves n , the number of moles.

Step 5 Solve the ideal gas law for the quantity to be determined (the unknown quantity). You may need to take a ratio of final states to initial states to eliminate the unknown quantities that are kept fixed.

Step 6 Substitute the known quantities, along with their units, into the appropriate equation, and obtain numerical solutions complete with units. Be certain to use absolute temperature and absolute pressure.

Step 7 Check the answer to see if it is reasonable: Does it make sense?

Exercise:

Check Your Understanding

Problem:

Liquids and solids have densities about 1000 times greater than gases. Explain how this implies that the distances between atoms and molecules in gases are about 10 times greater than the size of their atoms and molecules.

Solution:

Atoms and molecules are close together in solids and liquids. In gases they are separated by empty space. Thus gases have lower densities than liquids and solids. Density is mass per unit volume, and volume is related to the size of a body (such as a sphere) cubed. So if the distance between atoms and molecules increases by a factor of 10, then the volume occupied increases by a factor of 1000, and the density decreases by a factor of 1000.

Section Summary

- The ideal gas law relates the pressure and volume of a gas to the number of gas molecules and the temperature of the gas.
- The ideal gas law can be written in terms of the number of molecules of gas:

Equation:

$$PV = NkT,$$

where P is pressure, V is volume, T is temperature, N is number of molecules, and k is the Boltzmann constant

Equation:

$$k = 1.38 \times 10^{-23} \text{ J/K}.$$

- A mole is the number of atoms in a 12-g sample of carbon-12.
- The number of molecules in a mole is called Avogadro's number N_A ,

Equation:

$$N_A = 6.02 \times 10^{23} \text{ mol}^{-1}.$$

- A mole of any substance has a mass in grams equal to its molecular weight, which can be determined from the periodic table of elements.
- The ideal gas law can also be written and solved in terms of the number of moles of gas:

Equation:

$$PV = nRT,$$

where n is number of moles and R is the universal gas constant,

Equation:

$$R = 8.31 \text{ J/mol} \cdot \text{K}.$$

- The ideal gas law is generally valid at temperatures well above the boiling temperature.

Conceptual Questions

Exercise:

Problem:

Find out the human population of Earth. Is there a mole of people inhabiting Earth? If the average mass of a person is 60 kg, calculate the mass of a mole of people. How does the mass of a mole of people compare with the mass of Earth?

Exercise:

Problem:

Under what circumstances would you expect a gas to behave significantly differently than predicted by the ideal gas law?

Exercise:

Problem:

A constant-volume gas thermometer contains a fixed amount of gas. What property of the gas is measured to indicate its temperature?

Problems & Exercises

Exercise:

Problem:

The gauge pressure in your car tires is $2.50 \times 10^5 \text{ N/m}^2$ at a temperature of 35.0°C when you drive it onto a ferry boat to Alaska. What is their gauge pressure later, when their temperature has dropped to -40.0°C ?

Solution:

1.62 atm

Exercise:**Problem:**

Convert an absolute pressure of $7.00 \times 10^5 \text{ N/m}^2$ to gauge pressure in lb/in^2 . (This value was stated to be just less than 90.0 lb/in^2 in [\[link\]](#). Is it?)

Exercise:**Problem:**

Suppose a gas-filled incandescent light bulb is manufactured so that the gas inside the bulb is at atmospheric pressure when the bulb has a temperature of 20.0°C . (a) Find the gauge pressure inside such a bulb when it is hot, assuming its average temperature is 60.0°C (an approximation) and neglecting any change in volume due to thermal expansion or gas leaks. (b) The actual final pressure for the light bulb will be less than calculated in part (a) because the glass bulb will expand. What will the actual final pressure be, taking this into account? Is this a negligible difference?

Solution:

(a) 0.136 atm

(b) 0.135 atm. The difference between this value and the value from part (a) is negligible.

Exercise:**Problem:**

Large helium-filled balloons are used to lift scientific equipment to high altitudes. (a) What is the pressure inside such a balloon if it starts out at sea level with a temperature of 10.0°C and rises to an altitude where its volume is twenty times the original volume and its temperature is -50.0°C ? (b) What is the gauge pressure? (Assume atmospheric pressure is constant.)

Exercise:**Problem:**

Confirm that the units of nRT are those of energy for each value of R : (a) $8.31 \text{ J/mol} \cdot \text{K}$, (b) $1.99 \text{ cal/mol} \cdot \text{K}$, and (c) $0.0821 \text{ L} \cdot \text{atm/mol} \cdot \text{K}$.

Solution:

(a) $nRT = (\text{mol})(\text{J/mol} \cdot \text{K})(\text{K}) = \text{J}$

(b) $nRT = (\text{mol})(\text{cal/mol} \cdot \text{K})(\text{K}) = \text{cal}$

$$\begin{aligned}
 nRT &= (\text{mol})(\text{L} \cdot \text{atm}/\text{mol} \cdot \text{K})(\text{K}) \\
 \text{(c)} \quad &= \text{L} \cdot \text{atm} = (\text{m}^3)(\text{N}/\text{m}^2) \\
 &= \text{N} \cdot \text{m} = \text{J}
 \end{aligned}$$

Exercise:

Problem:

In the text, it was shown that $N/V = 2.68 \times 10^{25} \text{ m}^{-3}$ for gas at STP. (a) Show that this quantity is equivalent to $N/V = 2.68 \times 10^{19} \text{ cm}^{-3}$, as stated. (b) About how many atoms are there in one μm^3 (a cubic micrometer) at STP? (c) What does your answer to part (b) imply about the separation of atoms and molecules?

Exercise:

Problem:

Calculate the number of moles in the 2.00-L volume of air in the lungs of the average person. Note that the air is at 37.0°C (body temperature).

Solution:

$$7.86 \times 10^{-2} \text{ mol}$$

Exercise:

Problem:

An airplane passenger has 100 cm^3 of air in his stomach just before the plane takes off from a sea-level airport. What volume will the air have at cruising altitude if cabin pressure drops to $7.50 \times 10^4 \text{ N}/\text{m}^2$?

Exercise:

Problem:

(a) What is the volume (in km^3) of Avogadro's number of sand grains if each grain is a cube and has sides that are 1.0 mm long? (b) How many kilometers of beaches in length would this cover if the beach averages 100 m in width and 10.0 m in depth? Neglect air spaces between grains.

Solution:

$$\text{(a)} \quad 6.02 \times 10^5 \text{ km}^3$$

$$\text{(b)} \quad 6.02 \times 10^8 \text{ km}$$

Exercise:

Problem:

An expensive vacuum system can achieve a pressure as low as $1.00 \times 10^{-7} \text{ N/m}^2$ at 20°C . How many atoms are there in a cubic centimeter at this pressure and temperature?

Exercise:**Problem:**

The number density of gas atoms at a certain location in the space above our planet is about $1.00 \times 10^{11} \text{ m}^{-3}$, and the pressure is $2.75 \times 10^{-10} \text{ N/m}^2$ in this space. What is the temperature there?

Solution:

-73.9°C

Exercise:**Problem:**

A bicycle tire has a pressure of $7.00 \times 10^5 \text{ N/m}^2$ at a temperature of 18.0°C and contains 2.00 L of gas. What will its pressure be if you let out an amount of air that has a volume of 100 cm^3 at atmospheric pressure? Assume tire temperature and volume remain constant.

Exercise:**Problem:**

A high-pressure gas cylinder contains 50.0 L of toxic gas at a pressure of $1.40 \times 10^7 \text{ N/m}^2$ and a temperature of 25.0°C . Its valve leaks after the cylinder is dropped. The cylinder is cooled to dry ice temperature (-78.5°C) to reduce the leak rate and pressure so that it can be safely repaired. (a) What is the final pressure in the tank, assuming a negligible amount of gas leaks while being cooled and that there is no phase change? (b) What is the final pressure if one-tenth of the gas escapes? (c) To what temperature must the tank be cooled to reduce the pressure to 1.00 atm (assuming the gas does not change phase and that there is no leakage during cooling)? (d) Does cooling the tank appear to be a practical solution?

Solution:

(a) $9.14 \times 10^6 \text{ N/m}^2$

(b) $8.23 \times 10^6 \text{ N/m}^2$

(c) 2.16 K

(d) No. The final temperature needed is much too low to be easily achieved for a large object.

Exercise:

Problem:

Find the number of moles in 2.00 L of gas at 35.0°C and under $7.41 \times 10^7 \text{ N/m}^2$ of pressure.

Exercise:**Problem:**

Calculate the depth to which Avogadro's number of table tennis balls would cover Earth. Each ball has a diameter of 3.75 cm. Assume the space between balls adds an extra 25.0% to their volume and assume they are not crushed by their own weight.

Solution:

41 km

Exercise:**Problem:**

(a) What is the gauge pressure in a 25.0°C car tire containing 3.60 mol of gas in a 30.0 L volume? (b) What will its gauge pressure be if you add 1.00 L of gas originally at atmospheric pressure and 25.0°C? Assume the temperature returns to 25.0°C and the volume remains constant.

Exercise:**Problem:**

(a) In the deep space between galaxies, the density of atoms is as low as 10^6 atoms/m^3 , and the temperature is a frigid 2.7 K. What is the pressure? (b) What volume (in m^3) is occupied by 1 mol of gas? (c) If this volume is a cube, what is the length of its sides in kilometers?

Solution:

(a) $3.7 \times 10^{-17} \text{ Pa}$

(b) $6.0 \times 10^{17} \text{ m}^3$

(c) $8.4 \times 10^2 \text{ km}$

Glossary

ideal gas law

the physical law that relates the pressure and volume of a gas to the number of gas molecules or number of moles of gas and the temperature of the gas

Boltzmann constant

k , a physical constant that relates energy to temperature; $k = 1.38 \times 10^{-23} \text{ J/K}$

Avogadro's number

N_A , the number of molecules or atoms in one mole of a substance; $N_A = 6.02 \times 10^{23}$
particles/mole

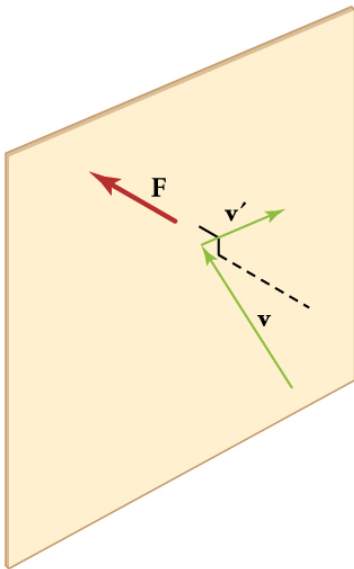
mole

the quantity of a substance whose mass (in grams) is equal to its molecular mass

Kinetic Theory: Atomic and Molecular Explanation of Pressure and Temperature

- Express the ideal gas law in terms of molecular mass and velocity.
- Define thermal energy.
- Calculate the kinetic energy of a gas molecule, given its temperature.
- Describe the relationship between the temperature of a gas and the kinetic energy of atoms and molecules.
- Describe the distribution of speeds of molecules in a gas.

We have developed macroscopic definitions of pressure and temperature. Pressure is the force divided by the area on which the force is exerted, and temperature is measured with a thermometer. We gain a better understanding of pressure and temperature from the kinetic theory of gases, which assumes that atoms and molecules are in continuous random motion.



When a molecule collides with a rigid wall, the component of its momentum perpendicular to the wall is reversed. A force is thus exerted on the wall, creating pressure.

[\[link\]](#) shows an elastic collision of a gas molecule with the wall of a container, so that it exerts a force on the wall (by Newton's third law). Because a huge number of molecules will collide with the wall in a short time, we observe an average force per unit area. These collisions are the source of pressure in a gas. As the number of molecules increases, the number of collisions and thus the pressure increase. Similarly, the gas pressure is higher if the average velocity of molecules is higher. The actual relationship is derived in the [Things Great and Small](#) feature below. The following relationship is found:

Equation:

$$PV = \frac{1}{3}Nm\overline{v^2},$$

where P is the pressure (average force per unit area), V is the volume of gas in the container, N is the number of molecules in the container, m is the mass of a molecule, and $\overline{v^2}$ is the average of the molecular speed squared.

What can we learn from this atomic and molecular version of the ideal gas law? We can derive a relationship between temperature and the average translational kinetic energy of molecules in a gas. Recall the previous expression of the ideal gas law:

Equation:

$$PV = NkT.$$

Equating the right-hand side of this equation with the right-hand side of $PV = \frac{1}{3}Nm\overline{v^2}$ gives

Equation:

$$\frac{1}{3}Nm\overline{v^2} = NkT.$$

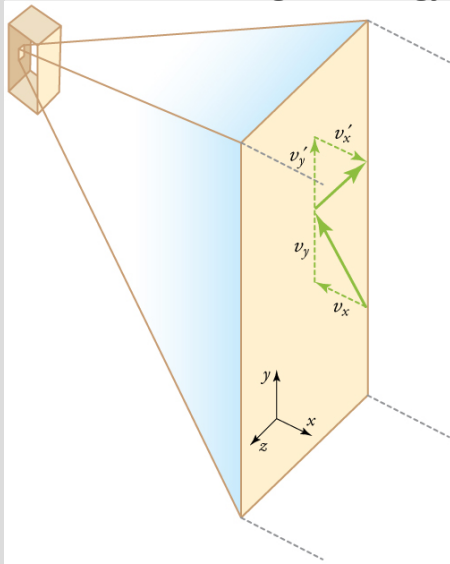
Note:

Making Connections: Things Great and Small—Atomic and Molecular Origin of Pressure in a Gas

[\[link\]](#) shows a box filled with a gas. We know from our previous discussions that putting more gas into the box produces greater pressure, and that increasing the temperature of the gas also produces a greater pressure. But why should increasing the temperature of the gas increase the pressure in the box? A look at the atomic and

molecular scale gives us some answers, and an alternative expression for the ideal gas law.

The figure shows an expanded view of an elastic collision of a gas molecule with the wall of a container. Calculating the average force exerted by such molecules will lead us to the ideal gas law, and to the connection between temperature and molecular kinetic energy. We assume that a molecule is small compared with the separation of molecules in the gas, and that its interaction with other molecules can be ignored. We also assume the wall is rigid and that the molecule's direction changes, but that its speed remains constant (and hence its kinetic energy and the magnitude of its momentum remain constant as well). This assumption is not always valid, but the same result is obtained with a more detailed description of the molecule's exchange of energy and momentum with the wall.



Gas in a box exerts an outward pressure on its walls. A molecule colliding with a rigid wall has the direction of its velocity and momentum in the x -direction reversed. This direction is perpendicular to the wall.

The components of its velocity momentum in the y - and z -directions are not changed, which means there is no force parallel to the wall.

If the molecule's velocity changes in the x -direction, its momentum changes from $-mv_x$ to $+mv_x$. Thus, its change in momentum is $\Delta mv = +mv_x - (-mv_x) = 2mv_x$. The force exerted on the molecule is given by

Equation:

$$F = \frac{\Delta p}{\Delta t} = \frac{2mv_x}{\Delta t}.$$

There is no force between the wall and the molecule until the molecule hits the wall. During the short time of the collision, the force between the molecule and wall is relatively large. We are looking for an average force; we take Δt to be the average time between collisions of the molecule with this wall. It is the time it would take the molecule to go across the box and back (a distance $2l$) at a speed of v_x . Thus $\Delta t = 2l/v_x$, and the expression for the force becomes

Equation:

$$F = \frac{2mv_x}{2l/v_x} = \frac{mv_x^2}{l}.$$

This force is due to *one* molecule. We multiply by the number of molecules N and use their average squared velocity to find the force

Equation:

$$F = N \frac{\overline{mv_x^2}}{l},$$

where the bar over a quantity means its average value. We would like to have the force in terms of the speed v , rather than the x -component of the velocity. We note that the total velocity squared is the sum of the squares of its components, so that

Equation:

$$\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}.$$

Because the velocities are random, their average components in all directions are the same:

Equation:

$$\overline{v_x^2} = \overline{v_y^2} = \overline{v_z^2}.$$

Thus,

Equation:

$$\overline{v^2} = 3\overline{v_x^2},$$

or

Equation:

$$\overline{v_x^2} = \frac{1}{3}\overline{v^2}.$$

Substituting $\frac{1}{3}\overline{v^2}$ into the expression for F gives

Equation:

$$F = N \frac{m\overline{v^2}}{3l}.$$

The pressure is F/A , so that we obtain

Equation:

$$P = \frac{F}{A} = N \frac{m\overline{v^2}}{3Al} = \frac{1}{3} \frac{Nm\overline{v^2}}{V},$$

where we used $V = Al$ for the volume. This gives the important result.

Equation:

$$PV = \frac{1}{3}Nm\overline{v^2}$$

This equation is another expression of the ideal gas law.

We can get the average kinetic energy of a molecule, $\frac{1}{2}mv^2$, from the right-hand side of the equation by canceling N and multiplying by $3/2$. This calculation produces the result that the average kinetic energy of a molecule is directly related to absolute temperature.

Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

The average translational kinetic energy of a molecule, $\overline{KE}$, is called **thermal energy**. The equation $\overline{KE} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$ is a molecular interpretation of temperature, and it has been found to be valid for gases and reasonably accurate in liquids and solids. It is another definition of temperature based on an expression of the molecular energy.

It is sometimes useful to rearrange $\overline{KE} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$, and solve for the average speed of molecules in a gas in terms of temperature,

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}},$$

where v_{rms} stands for root-mean-square (rms) speed.

Example:

Calculating Kinetic Energy and Speed of a Gas Molecule

(a) What is the average kinetic energy of a gas molecule at 20.0°C (room temperature)? (b) Find the rms speed of a nitrogen molecule (N_2) at this temperature.

Strategy for (a)

The known in the equation for the average kinetic energy is the temperature.

Equation:

$$\overline{KE} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

Before substituting values into this equation, we must convert the given temperature to kelvins. This conversion gives $T = (20.0 + 273) \text{ K} = 293 \text{ K}$.

Solution for (a)

The temperature alone is sufficient to find the average translational kinetic energy. Substituting the temperature into the translational kinetic energy equation gives

Equation:

$$\overline{KE} = \frac{3}{2}kT = \frac{3}{2}(1.38 \times 10^{-23} \text{ J/K})(293 \text{ K}) = 6.07 \times 10^{-21} \text{ J}.$$

Strategy for (b)

Finding the rms speed of a nitrogen molecule involves a straightforward calculation using the equation

Equation:

$$\sqrt{v^2} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}},$$

but we must first find the mass of a nitrogen molecule. Using the molecular mass of nitrogen N_2 from the periodic table,

Equation:

$$m = \frac{2(14.0067) \times 10^{-3} \text{ kg/mol}}{6.02 \times 10^{23} \text{ mol}^{-1}} = 4.65 \times 10^{-26} \text{ kg}.$$

Solution for (b)

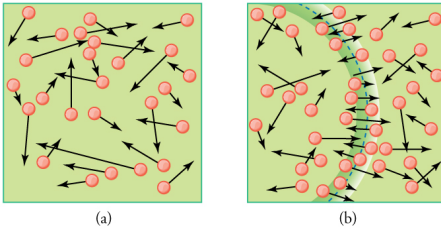
Substituting this mass and the value for k into the equation for v_{rms} yields

Equation:

$$v_{\text{rms}} = \sqrt{\frac{3kT}{m}} = \sqrt{\frac{3(1.38 \times 10^{-23} \text{ J/K})(293 \text{ K})}{4.65 \times 10^{-26} \text{ kg}}} = 511 \text{ m/s}.$$

Discussion

Note that the average kinetic energy of the molecule is independent of the type of molecule. The average translational kinetic energy depends only on absolute temperature. The kinetic energy is very small compared to macroscopic energies, so that we do not feel when an air molecule is hitting our skin. The rms velocity of the nitrogen molecule is surprisingly large. These large molecular velocities do not yield macroscopic movement of air, since the molecules move in all directions with equal likelihood. The *mean free path* (the distance a molecule can move on average between collisions) of molecules in air is very small, and so the molecules move rapidly but do not get very far in a second. The high value for rms speed is reflected in the speed of sound, however, which is about 340 m/s at room temperature. The faster the rms speed of air molecules, the faster that sound vibrations can be transferred through the air. The speed of sound increases with temperature and is greater in gases with small molecular masses, such as helium. (See [\[link\]](#).)



(a) There are many molecules moving so fast in an ordinary gas that they collide a billion times every second. (b) Individual molecules do not move very far in a small amount of time, but disturbances like sound waves are transmitted at speeds related to the molecular speeds.

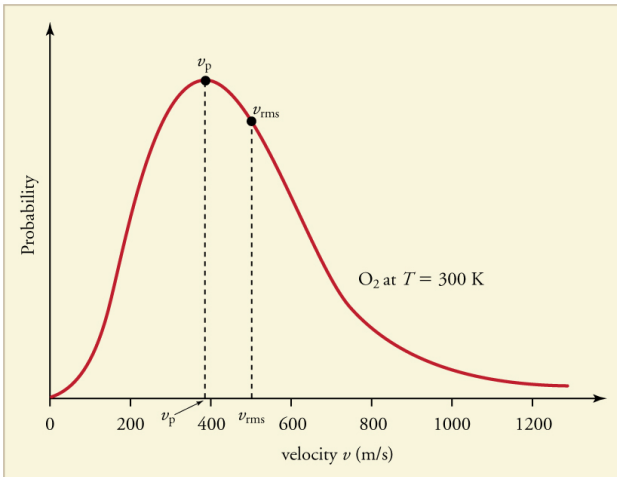
Note:

Making Connections: Historical Note—Kinetic Theory of Gases

The kinetic theory of gases was developed by Daniel Bernoulli (1700–1782), who is best known in physics for his work on fluid flow (hydrodynamics). Bernoulli's work predates the atomistic view of matter established by Dalton.

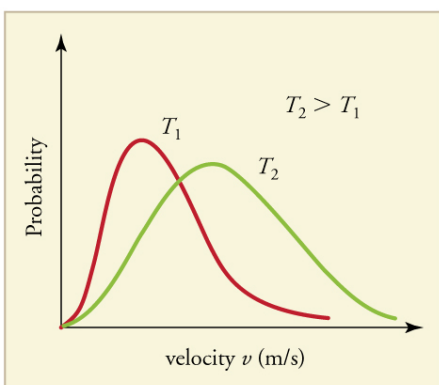
Distribution of Molecular Speeds

The motion of molecules in a gas is random in magnitude and direction for individual molecules, but a gas of many molecules has a predictable distribution of molecular speeds. This distribution is called the *Maxwell-Boltzmann distribution*, after its originators, who calculated it based on kinetic theory, and has since been confirmed experimentally. (See [\[link\]](#).) The distribution has a long tail, because a few molecules may go several times the rms speed. The most probable speed v_p is less than the rms speed v_{rms} . [\[link\]](#) shows that the curve is shifted to higher speeds at higher temperatures, with a broader range of speeds.



The Maxwell-Boltzmann distribution of molecular speeds in an ideal gas. The most likely speed v_p is less than the rms speed v_{rms} . Although very high speeds are possible, only a tiny fraction of the molecules have speeds that are an order of magnitude greater than v_{rms} .

The distribution of thermal speeds depends strongly on temperature. As temperature increases, the speeds are shifted to higher values and the distribution is broadened.



The Maxwell-Boltzmann distribution is shifted to

higher speeds and is broadened at higher temperatures.

What is the implication of the change in distribution with temperature shown in [\[link\]](#) for humans? All other things being equal, if a person has a fever, he or she is likely to lose more water molecules, particularly from linings along moist cavities such as the lungs and mouth, creating a dry sensation in the mouth.

Example:

Calculating Temperature: Escape Velocity of Helium Atoms

In order to escape Earth's gravity, an object near the top of the atmosphere (at an altitude of 100 km) must travel away from Earth at 11.1 km/s. This speed is called the *escape velocity*. At what temperature would helium atoms have an rms speed equal to the escape velocity?

Strategy

Identify the knowns and unknowns and determine which equations to use to solve the problem.

Solution

1. Identify the knowns: v is the escape velocity, 11.1 km/s.
2. Identify the unknowns: We need to solve for temperature, T . We also need to solve for the mass m of the helium atom.
3. Determine which equations are needed.

- To solve for mass m of the helium atom, we can use information from the periodic table:

Equation:

$$m = \frac{\text{molar mass}}{\text{number of atoms per mole}}.$$

- To solve for temperature T , we can rearrange either

Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

or

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}}$$

to yield

Equation:

$$T = \frac{m\overline{v^2}}{3k},$$

where k is the Boltzmann constant and m is the mass of a helium atom.

4. Plug the known values into the equations and solve for the unknowns.

Equation:

$$m = \frac{\text{molar mass}}{\text{number of atoms per mole}} = \frac{4.0026 \times 10^{-3} \text{ kg/mol}}{6.02 \times 10^{23} \text{ mol}} = 6.65 \times 10^{-27} \text{ kg}$$

Equation:

$$T = \frac{(6.65 \times 10^{-27} \text{ kg})(11.1 \times 10^3 \text{ m/s})^2}{3(1.38 \times 10^{-23} \text{ J/K})} = 1.98 \times 10^4 \text{ K}$$

Discussion

This temperature is much higher than atmospheric temperature, which is approximately 250 K (-25°C or -10°F) at high altitude. Very few helium atoms are left in the atmosphere, but there were many when the atmosphere was formed. The reason for the loss of helium atoms is that there are a small number of helium atoms with speeds higher than Earth's escape velocity even at normal temperatures. The speed of a helium atom changes from one instant to the next, so that at any instant, there is a small, but nonzero chance that the speed is greater than the escape speed and the molecule escapes from Earth's gravitational pull. Heavier molecules, such as oxygen, nitrogen, and water (very little of which reach a very high altitude), have smaller rms speeds, and so it is much less likely that any of them will have speeds greater than the escape velocity. In fact, so few have speeds above the escape velocity that billions of years are required to lose significant amounts of the atmosphere. [\[link\]](#) shows the impact of a lack of an atmosphere on the Moon. Because the gravitational pull of the Moon is much weaker, it has lost almost its

entire atmosphere. The comparison between Earth and the Moon is discussed in this chapter's Problems and Exercises.



This photograph of Apollo 17 Commander Eugene Cernan driving the lunar rover on the Moon in 1972 looks as though it was taken at night with a large spotlight. In fact, the light is coming from the Sun. Because the acceleration due to gravity on the Moon is so low (about $1/6$ that of Earth), the Moon's escape velocity is much smaller. As a result, gas molecules escape very easily from the Moon, leaving it with virtually no atmosphere. Even during the daytime, the sky is black because there is no gas to scatter sunlight. (credit: Harrison H. Schmitt/NASA)

Exercise:**Check Your Understanding****Problem:**

If you consider a very small object such as a grain of pollen, in a gas, then the number of atoms and molecules striking its surface would also be relatively small. Would the grain of pollen experience any fluctuations in pressure due to statistical fluctuations in the number of gas atoms and molecules striking it in a given amount of time?

Solution:

Yes. Such fluctuations actually occur for a body of any size in a gas, but since the numbers of atoms and molecules are immense for macroscopic bodies, the fluctuations are a tiny percentage of the number of collisions, and the averages spoken of in this section vary imperceptibly. Roughly speaking the fluctuations are proportional to the inverse square root of the number of collisions, so for small bodies they can become significant. This was actually observed in the 19th century for pollen grains in water, and is known as the Brownian effect.

Note:

PhET Explorations: Gas Properties

Pump gas molecules into a box and see what happens as you change the volume, add or remove heat, change gravity, and more. Measure the temperature and pressure, and discover how the properties of the gas vary in relation to each other.

[Click to open media in new browser.](#)

Section Summary

- Kinetic theory is the atomistic description of gases as well as liquids and solids.
- Kinetic theory models the properties of matter in terms of continuous random motion of atoms and molecules.
- The ideal gas law can also be expressed as

Equation:

$$PV = \frac{1}{3}Nm\overline{v^2},$$

where P is the pressure (average force per unit area), V is the volume of gas in the container, N is the number of molecules in the container, m is the mass of a molecule, and $\overline{v^2}$ is the average of the molecular speed squared.

- Thermal energy is defined to be the average translational kinetic energy $\overline{\text{KE}}$ of an atom or molecule.
- The temperature of gases is proportional to the average translational kinetic energy of atoms and molecules.

Equation:

$$\overline{\text{KE}} = \frac{1}{2}m\overline{v^2} = \frac{3}{2}kT$$

or

Equation:

$$\sqrt{\overline{v^2}} = v_{\text{rms}} = \sqrt{\frac{3kT}{m}}.$$

- The motion of individual molecules in a gas is random in magnitude and direction. However, a gas of many molecules has a predictable distribution of molecular speeds, known as the *Maxwell-Boltzmann distribution*.

Conceptual Questions

Exercise:

Problem:

How is momentum related to the pressure exerted by a gas? Explain on the atomic and molecular level, considering the behavior of atoms and molecules.

Problems & Exercises

Exercise:

Problem:

Some incandescent light bulbs are filled with argon gas. What is v_{rms} for argon atoms near the filament, assuming their temperature is 2500 K?

Solution:

$$1.25 \times 10^3 \text{ m/s}$$

Exercise:**Problem:**

Average atomic and molecular speeds (v_{rms}) are large, even at low temperatures. What is v_{rms} for helium atoms at 5.00 K, just one degree above helium's liquefaction temperature?

Exercise:**Problem:**

(a) What is the average kinetic energy in joules of hydrogen atoms on the 5500°C surface of the Sun? (b) What is the average kinetic energy of helium atoms in a region of the solar corona where the temperature is $6.00 \times 10^5 \text{ K}$?

Solution:

(a) $1.20 \times 10^{-19} \text{ J}$

(b) $1.24 \times 10^{-17} \text{ J}$

Exercise:**Problem:**

The escape velocity of any object from Earth is 11.2 km/s. (a) Express this speed in m/s and km/h. (b) At what temperature would oxygen molecules (molecular mass is equal to 32.0 g/mol) have an average velocity v_{rms} equal to Earth's escape velocity of 11.1 km/s?

Exercise:

Problem:

The escape velocity from the Moon is much smaller than from Earth and is only 2.38 km/s. At what temperature would hydrogen molecules (molecular mass is equal to 2.016 g/mol) have an average velocity v_{rms} equal to the Moon's escape velocity?

Solution:

458 K

Exercise:**Problem:**

Nuclear fusion, the energy source of the Sun, hydrogen bombs, and fusion reactors, occurs much more readily when the average kinetic energy of the atoms is high—that is, at high temperatures. Suppose you want the atoms in your fusion experiment to have average kinetic energies of 6.40×10^{-14} J. What temperature is needed?

Exercise:**Problem:**

Suppose that the average velocity (v_{rms}) of carbon dioxide molecules (molecular mass is equal to 44.0 g/mol) in a flame is found to be 1.05×10^5 m/s. What temperature does this represent?

Solution:

1.95×10^7 K

Exercise:**Problem:**

Hydrogen molecules (molecular mass is equal to 2.016 g/mol) have an average velocity v_{rms} equal to 193 m/s. What is the temperature?

Exercise:

Problem:

Much of the gas near the Sun is atomic hydrogen. Its temperature would have to be 1.5×10^7 K for the average velocity v_{rms} to equal the escape velocity from the Sun. What is that velocity?

Solution:

$$6.09 \times 10^5 \text{ m/s}$$

Exercise:**Problem:**

There are two important isotopes of uranium— ^{235}U and ^{238}U ; these isotopes are nearly identical chemically but have different atomic masses. Only ^{235}U is very useful in nuclear reactors. One of the techniques for separating them (gas diffusion) is based on the different average velocities v_{rms} of uranium hexafluoride gas, UF_6 . (a) The molecular masses for $^{235}\text{U UF}_6$ and $^{238}\text{U UF}_6$ are 349.0 g/mol and 352.0 g/mol, respectively. What is the ratio of their average velocities? (b) At what temperature would their average velocities differ by 1.00 m/s? (c) Do your answers in this problem imply that this technique may be difficult?

Glossary

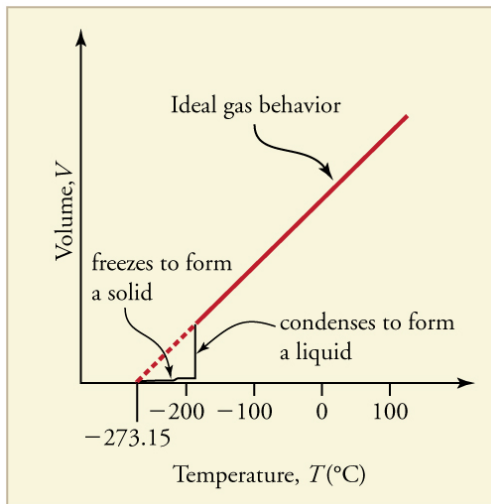
thermal energy

$\overline{\text{KE}}$, the average translational kinetic energy of a molecule

Phase Changes

- Interpret a phase diagram.
- State Dalton's law.
- Identify and describe the triple point of a gas from its phase diagram.
- Describe the state of equilibrium between a liquid and a gas, a liquid and a solid, and a gas and a solid.

Up to now, we have considered the behavior of ideal gases. Real gases are like ideal gases at high temperatures. At lower temperatures, however, the interactions between the molecules and their volumes cannot be ignored. The molecules are very close (condensation occurs) and there is a dramatic decrease in volume, as seen in [\[link\]](#). The substance changes from a gas to a liquid. When a liquid is cooled to even lower temperatures, it becomes a solid. The volume never reaches zero because of the finite volume of the molecules.



A sketch of volume versus temperature for a real gas at constant pressure. The linear (straight line) part of the graph represents ideal gas behavior—volume and temperature are directly and positively related and

the line extrapolates to zero volume at -273.15°C , or absolute zero. When the gas becomes a liquid, however, the volume actually decreases precipitously at the liquefaction point. The volume decreases slightly once the substance is solid, but it never becomes zero.

High pressure may also cause a gas to change phase to a liquid. Carbon dioxide, for example, is a gas at room temperature and atmospheric pressure, but becomes a liquid under sufficiently high pressure. If the pressure is reduced, the temperature drops and the liquid carbon dioxide solidifies into a snow-like substance at the temperature -78°C . Solid CO_2 is called “dry ice.” Another example of a gas that can be in a liquid phase is liquid nitrogen (LN_2). LN_2 is made by liquefaction of atmospheric air (through compression and cooling). It boils at 77 K (-196°C) at atmospheric pressure. LN_2 is useful as a refrigerant and allows for the preservation of blood, sperm, and other biological materials. It is also used to reduce noise in electronic sensors and equipment, and to help cool down their current-carrying wires. In dermatology, LN_2 is used to freeze and painlessly remove warts and other growths from the skin.

PV Diagrams

We can examine aspects of the behavior of a substance by plotting a graph of pressure versus volume, called a **PV diagram**. When the substance behaves like an ideal gas, the ideal gas law describes the relationship between its pressure and volume. That is,

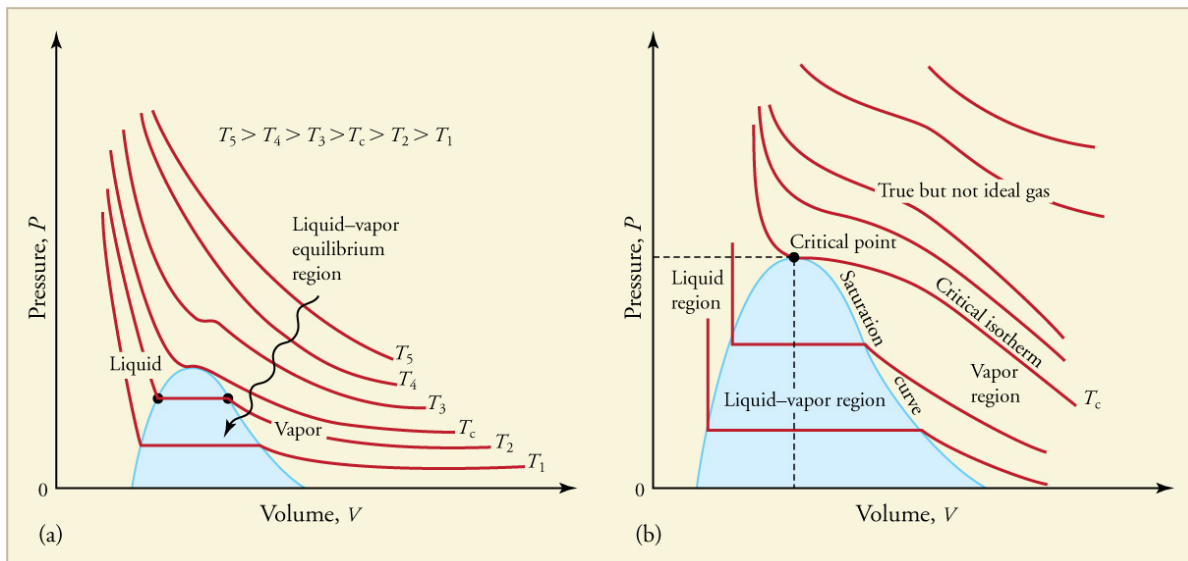
Equation:

$$PV = NkT \text{ (ideal gas).}$$

Now, assuming the number of molecules and the temperature are fixed,
Equation:

$$PV = \text{constant (ideal gas, constant temperature).}$$

For example, the volume of the gas will decrease as the pressure increases. If you plot the relationship $PV = \text{constant}$ on a PV diagram, you find a hyperbola. [\[link\]](#) shows a graph of pressure versus volume. The hyperbolas represent ideal-gas behavior at various fixed temperatures, and are called *isotherms*. At lower temperatures, the curves begin to look less like hyperbolas—the gas is not behaving ideally and may even contain liquid. There is a **critical point**—that is, a **critical temperature**—above which liquid cannot exist. At sufficiently high pressure above the critical point, the gas will have the density of a liquid but will not condense. Carbon dioxide, for example, cannot be liquefied at a temperature above 31.0°C. **Critical pressure** is the minimum pressure needed for liquid to exist at the critical temperature. [\[link\]](#) lists representative critical temperatures and pressures.



PV diagrams. (a) Each curve (isotherm) represents the relationship

between P and V at a fixed temperature; the upper curves are at higher temperatures. The lower curves are not hyperbolas, because the gas is no longer an ideal gas. (b) An expanded portion of the PV diagram for low temperatures, where the phase can change from a gas to a liquid. The term “vapor” refers to the gas phase when it exists at a temperature below the boiling temperature.

Substance	Critical temperature		Critical pressure	
	K	°C	Pa	atm
Water	647.4	374.3	22.12×10^6	219.0
Sulfur dioxide	430.7	157.6	7.88×10^6	78.0
Ammonia	405.5	132.4	11.28×10^6	111.7
Carbon dioxide	304.2	31.1	7.39×10^6	73.2

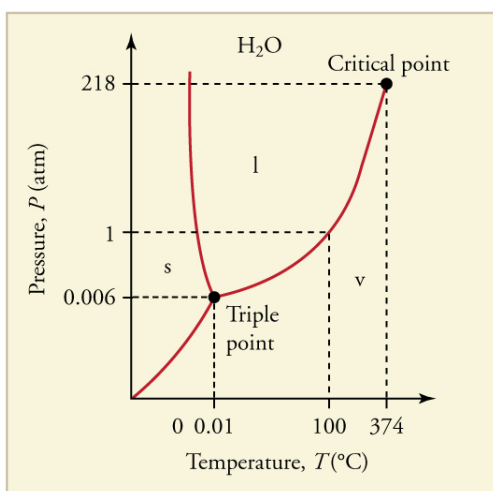
Substance	Critical temperature		Critical pressure	
	K	°C	Pa	atm
Oxygen	154.8	−118.4	5.08×10^6	50.3
Nitrogen	126.2	−146.9	3.39×10^6	33.6
Hydrogen	33.3	−239.9	1.30×10^6	12.9
Helium	5.3	−267.9	0.229×10^6	2.27

Critical Temperatures and Pressures

Phase Diagrams

The plots of pressure versus temperatures provide considerable insight into thermal properties of substances. There are well-defined regions on these graphs that correspond to various phases of matter, so PT graphs are called **phase diagrams**. [\[link\]](#) shows the phase diagram for water. Using the graph, if you know the pressure and temperature you can determine the phase of water. The solid lines—boundaries between phases—indicate temperatures and pressures at which the phases coexist (that is, they exist together in ratios, depending on pressure and temperature). For example, the boiling point of water is 100°C at 1.00 atm. As the pressure increases, the boiling temperature rises steadily to 374°C at a pressure of 218 atm. A pressure cooker (or even a covered pot) will cook food faster because the

water can exist as a liquid at temperatures greater than 100°C without all boiling away. The curve ends at a point called the *critical point*, because at higher temperatures the liquid phase does not exist at any pressure. The critical point occurs at the critical temperature, as you can see for water from [\[link\]](#). The critical temperature for oxygen is -118°C , so oxygen cannot be liquefied above this temperature.



The phase diagram (PT graph) for water. Note that the axes are nonlinear and the graph is not to scale. This graph is simplified—there are several other exotic phases of ice at higher pressures.

Similarly, the curve between the solid and liquid regions in [\[link\]](#) gives the melting temperature at various pressures. For example, the melting point is 0°C at 1.00 atm, as expected. Note that, at a fixed temperature, you can change the phase from solid (ice) to liquid (water) by increasing the pressure. Ice melts from pressure in the hands of a snowball maker. From

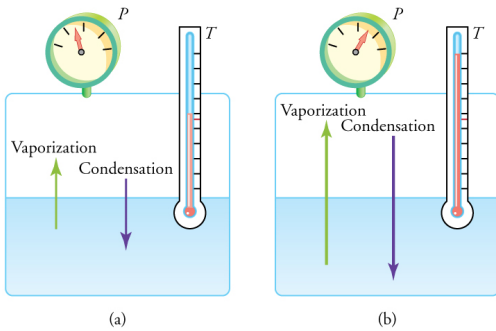
the phase diagram, we can also say that the melting temperature of ice falls with increased pressure. When a car is driven over snow, the increased pressure from the tires melts the snowflakes; afterwards the water refreezes and forms an ice layer.

At sufficiently low pressures there is no liquid phase, but the substance can exist as either gas or solid. For water, there is no liquid phase at pressures below 0.00600 atm. The phase change from solid to gas is called **sublimation**. It accounts for large losses of snow pack that never make it into a river, the routine automatic defrosting of a freezer, and the freeze-drying process applied to many foods. Carbon dioxide, on the other hand, sublimates at standard atmospheric pressure of 1 atm. (The solid form of CO_2 is known as dry ice because it does not melt. Instead, it moves directly from the solid to the gas state.)

All three curves on the phase diagram meet at a single point, the **triple point**, where all three phases exist in equilibrium. For water, the triple point occurs at 273.16 K (0.01°C), and is a more accurate calibration temperature than the melting point of water at 1.00 atm, or 273.15 K (0.0°C). See [\[link\]](#) for the triple point values of other substances.

Equilibrium

Liquid and gas phases are in equilibrium at the boiling temperature. (See [\[link\]](#).) If a substance is in a closed container at the boiling point, then the liquid is boiling and the gas is condensing at the same rate without net change in their relative amount. Molecules in the liquid escape as a gas at the same rate at which gas molecules stick to the liquid, or form droplets and become part of the liquid phase. The combination of temperature and pressure has to be “just right”; if the temperature and pressure are increased, equilibrium is maintained by the same increase of boiling and condensation rates.



Equilibrium between liquid and gas at two different boiling points inside a closed container.

(a) The rates of boiling and condensation are equal at this combination of temperature and pressure, so the liquid and gas phases are in equilibrium. (b) At a higher temperature, the boiling rate is faster and the rates at which molecules leave the liquid and enter the gas are also faster. Because there are more molecules in the gas, the gas pressure is higher and the rate at which gas molecules condense and enter the liquid is faster. As a result the gas and liquid are in equilibrium at this higher temperature.

Substance	Temperature		Pressure	
	K	°C	Pa	atm
Water	273.16	0.01	6.10×10^2	0.00600
Carbon dioxide	216.55	-56.60	5.16×10^5	5.11
Sulfur dioxide	197.68	-75.47	1.67×10^3	0.0167
Ammonia	195.40	-77.75	6.06×10^3	0.0600
Nitrogen	63.18	-210.0	1.25×10^4	0.124
Oxygen	54.36	-218.8	1.52×10^2	0.00151
Hydrogen	13.84	-259.3	7.04×10^3	0.0697

Triple Point Temperatures and Pressures

One example of equilibrium between liquid and gas is that of water and steam at 100°C and 1.00 atm. This temperature is the boiling point at that pressure, so they should exist in equilibrium. Why does an open pot of water at 100°C boil completely away? The gas surrounding an open pot is

not pure water: it is mixed with air. If pure water and steam are in a closed container at 100°C and 1.00 atm , they would coexist—but with air over the pot, there are fewer water molecules to condense, and water boils. What about water at 20.0°C and 1.00 atm ? This temperature and pressure correspond to the liquid region, yet an open glass of water at this temperature will completely evaporate. Again, the gas around it is air and not pure water vapor, so that the reduced evaporation rate is greater than the condensation rate of water from dry air. If the glass is sealed, then the liquid phase remains. We call the gas phase a **vapor** when it exists, as it does for water at 20.0°C , at a temperature below the boiling temperature.

Exercise:

Check Your Understanding

Problem:

Explain why a cup of water (or soda) with ice cubes stays at 0°C , even on a hot summer day.

Solution:

The ice and liquid water are in thermal equilibrium, so that the temperature stays at the freezing temperature as long as ice remains in the liquid. (Once all of the ice melts, the water temperature will start to rise.)

Vapor Pressure, Partial Pressure, and Dalton's Law

Vapor pressure is defined as the pressure at which a gas coexists with its solid or liquid phase. Vapor pressure is created by faster molecules that break away from the liquid or solid and enter the gas phase. The vapor pressure of a substance depends on both the substance and its temperature—an increase in temperature increases the vapor pressure.

Partial pressure is defined as the pressure a gas would create if it occupied the total volume available. In a mixture of gases, *the total pressure is the sum of partial pressures of the component gases*, assuming ideal gas behavior and no chemical reactions between the components. This law is

known as **Dalton's law of partial pressures**, after the English scientist John Dalton (1766–1844), who proposed it. Dalton's law is based on kinetic theory, where each gas creates its pressure by molecular collisions, independent of other gases present. It is consistent with the fact that pressures add according to [Pascal's Principle](#). Thus water evaporates and ice sublimates when their vapor pressures exceed the partial pressure of water vapor in the surrounding mixture of gases. If their vapor pressures are less than the partial pressure of water vapor in the surrounding gas, liquid droplets or ice crystals (frost) form.

Exercise:

Check Your Understanding

Problem:

Is energy transfer involved in a phase change? If so, will energy have to be supplied to change phase from solid to liquid and liquid to gas? What about gas to liquid and liquid to solid? Why do they spray the orange trees with water in Florida when the temperatures are near or just below freezing?

Solution:

Yes, energy transfer is involved in a phase change. We know that atoms and molecules in solids and liquids are bound to each other because we know that force is required to separate them. So in a phase change from solid to liquid and liquid to gas, a force must be exerted, perhaps by collision, to separate atoms and molecules. Force exerted through a distance is work, and energy is needed to do work to go from solid to liquid and liquid to gas. This is intuitively consistent with the need for energy to melt ice or boil water. The converse is also true. Going from gas to liquid or liquid to solid involves atoms and molecules pushing together, doing work and releasing energy.

Note:

PhET Explorations: States of Matter—Basics

Heat, cool, and compress atoms and molecules and watch as they change between solid, liquid, and gas phases.

https://phet.colorado.edu/sims/html/states-of-matter-basics/latest/states-of-matter-basics_en.html

Section Summary

- Most substances have three distinct phases: gas, liquid, and solid.
- Phase changes among the various phases of matter depend on temperature and pressure.
- The existence of the three phases with respect to pressure and temperature can be described in a phase diagram.
- Two phases coexist (i.e., they are in thermal equilibrium) at a set of pressures and temperatures. These are described as a line on a phase diagram.
- The three phases coexist at a single pressure and temperature. This is known as the triple point and is described by a single point on a phase diagram.
- A gas at a temperature below its boiling point is called a vapor.
- Vapor pressure is the pressure at which a gas coexists with its solid or liquid phase.
- Partial pressure is the pressure a gas would create if it existed alone.
- Dalton's law states that the total pressure is the sum of the partial pressures of all of the gases present.

Conceptual Questions

Exercise:

Problem:

A pressure cooker contains water and steam in equilibrium at a pressure greater than atmospheric pressure. How does this greater pressure increase cooking speed?

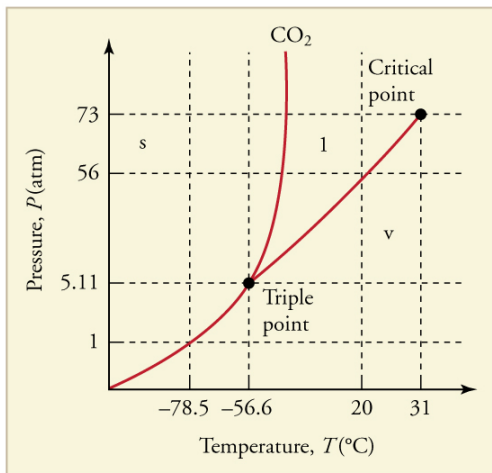
Exercise:

Problem:

Why does condensation form most rapidly on the coldest object in a room—for example, on a glass of ice water?

Exercise:**Problem:**

What is the vapor pressure of solid carbon dioxide (dry ice) at -78.5°C ?



The phase diagram for carbon dioxide. The axes are nonlinear, and the graph is not to scale. Dry ice is solid carbon dioxide and has a sublimation temperature of -78.5°C .

Exercise:**Problem:**

Can carbon dioxide be liquefied at room temperature (20°C)? If so, how? If not, why not? (See [\[link\]](#).)

Exercise:**Problem:**

Oxygen cannot be liquefied at room temperature by placing it under a large enough pressure to force its molecules together. Explain why this is.

Exercise:

Problem: What is the distinction between gas and vapor?

Glossary

PV diagram

a graph of pressure vs. volume

critical point

the temperature above which a liquid cannot exist

critical temperature

the temperature above which a liquid cannot exist

critical pressure

the minimum pressure needed for a liquid to exist at the critical temperature

vapor

a gas at a temperature below the boiling temperature

vapor pressure

the pressure at which a gas coexists with its solid or liquid phase

phase diagram

a graph of pressure vs. temperature of a particular substance, showing at which pressures and temperatures the three phases of the substance occur

triple point

the pressure and temperature at which a substance exists in equilibrium as a solid, liquid, and gas

sublimation

the phase change from solid to gas

partial pressure

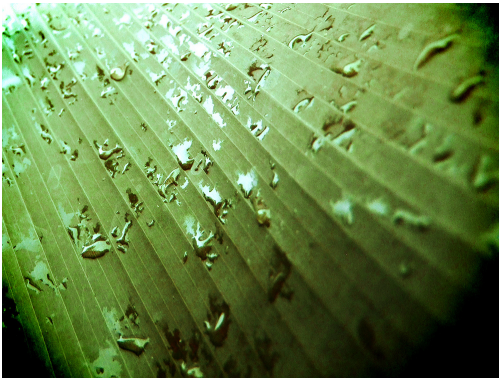
the pressure a gas would create if it occupied the total volume of space available

Dalton's law of partial pressures

the physical law that states that the total pressure of a gas is the sum of partial pressures of the component gases

Humidity, Evaporation, and Boiling

- Explain the relationship between vapor pressure of water and the capacity of air to hold water vapor.
- Explain the relationship between relative humidity and partial pressure of water vapor in the air.
- Calculate vapor density using vapor pressure.
- Calculate humidity and dew point.



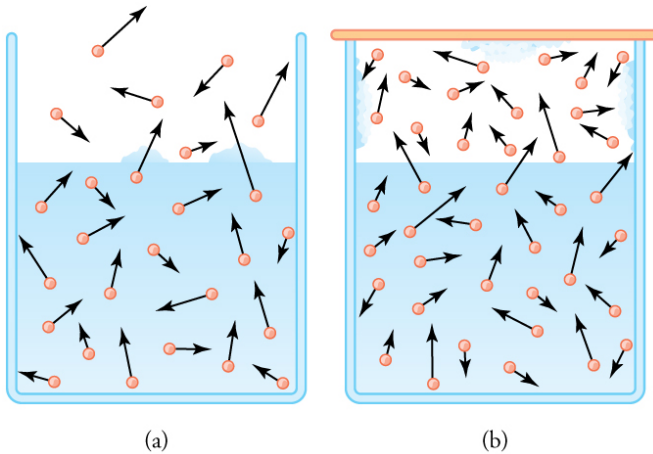
Dew drops like these, on a banana leaf photographed just after sunrise, form when the air temperature drops to or below the dew point. At the dew point, the rate at which water molecules join together is greater than the rate at which they separate, and some of the water condenses to form droplets. (credit: Aaron Escobar, Flickr)

The expression “it’s not the heat, it’s the humidity” makes a valid point. We keep cool in hot weather by evaporating sweat from our skin and water

from our breathing passages. Because evaporation is inhibited by high humidity, we feel hotter at a given temperature when the humidity is high. Low humidity, on the other hand, can cause discomfort from excessive drying of mucous membranes and can lead to an increased risk of respiratory infections.

When we say humidity, we really mean **relative humidity**. Relative humidity tells us how much water vapor is in the air compared with the maximum possible. At its maximum, denoted as **saturation**, the relative humidity is 100%, and evaporation is inhibited. The amount of water vapor in the air depends on temperature. For example, relative humidity rises in the evening, as air temperature declines, sometimes reaching the **dew point**. At the dew point temperature, relative humidity is 100%, and fog may result from the condensation of water droplets if they are small enough to stay in suspension. Conversely, if you wish to dry something (perhaps your hair), it is more effective to blow hot air over it rather than cold air, because, among other things, the increase in temperature increases the energy of the molecules, so the rate of evaporation increases.

The amount of water vapor in the air depends on the vapor pressure of water. The liquid and solid phases are continuously giving off vapor because some of the molecules have high enough speeds to enter the gas phase; see [\[link\]](#)(a). If a lid is placed over the container, as in [\[link\]](#)(b), evaporation continues, increasing the pressure, until sufficient vapor has built up for condensation to balance evaporation. Then equilibrium has been achieved, and the vapor pressure is equal to the partial pressure of water in the container. Vapor pressure increases with temperature because molecular speeds are higher as temperature increases. [\[link\]](#) gives representative values of water vapor pressure over a range of temperatures.



(a) Because of the distribution of speeds and kinetic energies, some water molecules can break away to the vapor phase even at temperatures below the ordinary boiling point. (b) If the container is sealed, evaporation will continue until there is enough vapor density for the condensation rate to equal the evaporation rate. This vapor density and the partial pressure it creates are the saturation values. They increase with temperature and are independent of the presence of other gases, such as air. They depend only on the vapor pressure of water.

Relative humidity is related to the partial pressure of water vapor in the air. At 100% humidity, the partial pressure is equal to the vapor pressure, and no more water can enter the vapor phase. If the partial pressure is less than the vapor pressure, then evaporation will take place, as humidity is less than 100%. If the partial pressure is greater than the vapor pressure, condensation takes place. In everyday language, people sometimes refer to

the capacity of air to “hold” water vapor, but this is not actually what happens. The water vapor is not held by the air. The amount of water in air is determined by the vapor pressure of water and has nothing to do with the properties of air.

Temperature (°C)	Vapor pressure (Pa)	Saturation vapor density (g/m ³)
−50	4.0	0.039
−20	1.04×10^2	0.89
−10	2.60×10^2	2.36
0	6.10×10^2	4.84
5	8.68×10^2	6.80
10	1.19×10^3	9.40

Temperature (°C)	Vapor pressure (Pa)	Saturation vapor density (g/m ³)
15	1.69×10^3	12.8
20	2.33×10^3	17.2
25	3.17×10^3	23.0
30	4.24×10^3	30.4
37	6.31×10^3	44.0
40	7.34×10^3	51.1
50	1.23×10^4	82.4
60	1.99×10^4	130
70	3.12×10^4	197

Temperature (°C)	Vapor pressure (Pa)	Saturation vapor density (g/m ³)
80	4.73×10^4	294
90	7.01×10^4	418
95	8.59×10^4	505
100	1.01×10^5	598
120	1.99×10^5	1095
150	4.76×10^5	2430
200	1.55×10^6	7090
220	2.32×10^6	10,200

Saturation Vapor Density of Water

Example:**Calculating Density Using Vapor Pressure**

[\[link\]](#) gives the vapor pressure of water at 20.0°C as 2.33×10^3 Pa. Use the ideal gas law to calculate the density of water vapor in g/m³ that would create a partial pressure equal to this vapor pressure. Compare the result with the saturation vapor density given in the table.

Strategy

To solve this problem, we need to break it down into a two steps. The partial pressure follows the ideal gas law,

Equation:

$$PV = nRT,$$

where n is the number of moles. If we solve this equation for n/V to calculate the number of moles per cubic meter, we can then convert this quantity to grams per cubic meter as requested. To do this, we need to use the molecular mass of water, which is given in the periodic table.

Solution

1. Identify the knowns and convert them to the proper units:

- a. temperature $T = 20^\circ\text{C} = 293$ K
- b. vapor pressure P of water at 20°C is 2.33×10^3 Pa
- c. molecular mass of water is 18.0 g/mol

2. Solve the ideal gas law for n/V .

Equation:

$$\frac{n}{V} = \frac{P}{RT}$$

3. Substitute known values into the equation and solve for n/V .

Equation:

$$\frac{n}{V} = \frac{P}{RT} = \frac{2.33 \times 10^3 \text{ Pa}}{(8.31 \text{ J/mol} \cdot \text{K})(293 \text{ K})} = 0.957 \text{ mol/m}^3$$

4. Convert the density in moles per cubic meter to grams per cubic meter.

Equation:

$$\rho = \left(0.957 \frac{\text{mol}}{\text{m}^3}\right) \left(\frac{18.0 \text{ g}}{\text{mol}}\right) = 17.2 \text{ g/m}^3$$

Discussion

The density is obtained by assuming a pressure equal to the vapor pressure of water at 20.0°C. The density found is identical to the value in [\[link\]](#), which means that a vapor density of 17.2 g/m³ at 20.0°C creates a partial pressure of 2.33×10^3 Pa, equal to the vapor pressure of water at that temperature. If the partial pressure is equal to the vapor pressure, then the liquid and vapor phases are in equilibrium, and the relative humidity is 100%. Thus, there can be no more than 17.2 g of water vapor per m³ at 20.0°C, so that this value is the saturation vapor density at that temperature. This example illustrates how water vapor behaves like an ideal gas: the pressure and density are consistent with the ideal gas law (assuming the density in the table is correct). The saturation vapor densities listed in [\[link\]](#) are the maximum amounts of water vapor that air can hold at various temperatures.

Note:**Percent Relative Humidity**

We define **percent relative humidity** as the ratio of vapor density to saturation vapor density, or

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100$$

We can use this and the data in [\[link\]](#) to do a variety of interesting calculations, keeping in mind that relative humidity is based on the comparison of the partial pressure of water vapor in air and ice.

Example:**Calculating Humidity and Dew Point**

(a) Calculate the percent relative humidity on a day when the temperature is 25.0°C and the air contains 9.40 g of water vapor per m³. (b) At what temperature will this air reach 100% relative humidity (the saturation density)? This temperature is the dew point. (c) What is the humidity when the air temperature is 25.0°C and the dew point is – 10.0°C?

Strategy and Solution

(a) Percent relative humidity is defined as the ratio of vapor density to saturation vapor density.

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100$$

The first is given to be 9.40 g/m³, and the second is found in [\[link\]](#) to be 23.0 g/m³. Thus,

Equation:

$$\text{percent relative humidity} = \frac{9.40 \text{ g/m}^3}{23.0 \text{ g/m}^3} \times 100 = 40.9\%$$

(b) The air contains 9.40 g/m³ of water vapor. The relative humidity will be 100% at a temperature where 9.40 g/m³ is the saturation density. Inspection of [\[link\]](#) reveals this to be the case at 10.0°C, where the relative humidity will be 100%. That temperature is called the dew point for air with this concentration of water vapor.

(c) Here, the dew point temperature is given to be – 10.0°C. Using [\[link\]](#), we see that the vapor density is 2.36 g/m³, because this value is the saturation vapor density at – 10.0°C. The saturation vapor density at 25.0°C is seen to be 23.0 g/m³. Thus, the relative humidity at 25.0°C is

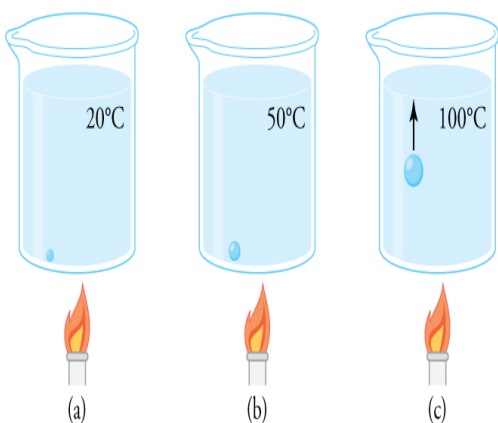
Equation:

$$\text{percent relative humidity} = \frac{2.36 \text{ g/m}^3}{23.0 \text{ g/m}^3} \times 100 = 10.3\%.$$

Discussion

The importance of dew point is that air temperature cannot drop below 10.0°C in part (b), or -10.0°C in part (c), without water vapor condensing out of the air. If condensation occurs, considerable transfer of heat occurs (discussed in [Heat and Heat Transfer Methods](#)), which prevents the temperature from further dropping. When dew points are below 0°C , freezing temperatures are a greater possibility, which explains why farmers keep track of the dew point. Low humidity in deserts means low dew-point temperatures. Thus condensation is unlikely. If the temperature drops, vapor does not condense in liquid drops. Because no heat is released into the air, the air temperature drops more rapidly compared to air with higher humidity. Likewise, at high temperatures, liquid droplets do not evaporate, so that no heat is removed from the gas to the liquid phase. This explains the large range of temperature in arid regions.

Why does water boil at 100°C ? You will note from [\[link\]](#) that the vapor pressure of water at 100°C is $1.01 \times 10^5 \text{ Pa}$, or 1.00 atm. Thus, it can evaporate without limit at this temperature and pressure. But why does it form bubbles when it boils? This is because water ordinarily contains significant amounts of dissolved air and other impurities, which are observed as small bubbles of air in a glass of water. If a bubble starts out at the bottom of the container at 20°C , it contains water vapor (about 2.30%). The pressure inside the bubble is fixed at 1.00 atm (we ignore the slight pressure exerted by the water around it). As the temperature rises, the amount of air in the bubble stays the same, but the water vapor increases; the bubble expands to keep the pressure at 1.00 atm. At 100°C , water vapor enters the bubble continuously since the partial pressure of water is equal to 1.00 atm in equilibrium. It cannot reach this pressure, however, since the bubble also contains air and total pressure is 1.00 atm. The bubble grows in size and thereby increases the buoyant force. The bubble breaks away and rises rapidly to the surface—we call this boiling! (See [\[link\]](#).)



- (a) An air bubble in water starts out saturated with water vapor at 20°C. (b) As the temperature rises, water vapor enters the bubble because its vapor pressure increases. The bubble expands to keep its pressure at 1.00 atm. (c) At 100°C, water vapor enters the bubble continuously because water's vapor pressure exceeds its partial pressure in the bubble, which must be less than 1.00 atm. The bubble grows and rises to the surface.

Exercise:
Check Your Understanding

Problem:

Freeze drying is a process in which substances, such as foods, are dried by placing them in a vacuum chamber and lowering the atmospheric pressure around them. How does the lowered atmospheric pressure speed the drying process, and why does it cause the temperature of the food to drop?

Solution:

Decreased the atmospheric pressure results in decreased partial pressure of water, hence a lower humidity. So evaporation of water from food, for example, will be enhanced. The molecules of water most likely to break away from the food will be those with the greatest velocities. Those remaining thus have a lower average velocity and a lower temperature. This can (and does) result in the freezing and drying of the food; hence the process is aptly named freeze drying.

Note:

PhET Explorations: States of Matter

Watch different types of molecules form a solid, liquid, or gas. Add or remove heat and watch the phase change. Change the temperature or volume of a container and see a pressure-temperature diagram respond in real time. Relate the interaction potential to the forces between molecules.

https://phet.colorado.edu/sims/html/states-of-matter/latest/states-of-matter_en.html

Section Summary

- Relative humidity is the fraction of water vapor in a gas compared to the saturation value.
- The saturation vapor density can be determined from the vapor pressure for a given temperature.

- Percent relative humidity is defined to be

Equation:

$$\text{percent relative humidity} = \frac{\text{vapor density}}{\text{saturation vapor density}} \times 100.$$

- The dew point is the temperature at which air reaches 100% relative humidity.

Conceptual Questions

Exercise:

Problem:

Because humidity depends only on water's vapor pressure and temperature, are the saturation vapor densities listed in [\[link\]](#) valid in an atmosphere of helium at a pressure of $1.01 \times 10^5 \text{ N/m}^2$, rather than air? Are those values affected by altitude on Earth?

Exercise:

Problem:

Why does a beaker of 40.0°C water placed in a vacuum chamber start to boil as the chamber is evacuated (air is pumped out of the chamber)? At what pressure does the boiling begin? Would food cook any faster in such a beaker?

Exercise:

Problem:

Why does rubbing alcohol evaporate much more rapidly than water at STP (standard temperature and pressure)?

Problems & Exercises

Exercise:

Problem:

Dry air is 78.1% nitrogen. What is the partial pressure of nitrogen when the atmospheric pressure is $1.01 \times 10^5 \text{ N/m}^2$?

Solution:

$$7.89 \times 10^4 \text{ Pa}$$

Exercise:**Problem:**

(a) What is the vapor pressure of water at 20.0°C? (b) What percentage of atmospheric pressure does this correspond to? (c) What percent of 20.0°C air is water vapor if it has 100% relative humidity? (The density of dry air at 20.0°C is 1.20 kg/m^3 .)

Exercise:**Problem:**

Pressure cookers increase cooking speed by raising the boiling temperature of water above its value at atmospheric pressure. (a) What pressure is necessary to raise the boiling point to 120.0°C? (b) What gauge pressure does this correspond to?

Solution:

(a) $1.99 \times 10^5 \text{ Pa}$

(b) 0.97 atm

Exercise:

Problem:

(a) At what temperature does water boil at an altitude of 1500 m (about 5000 ft) on a day when atmospheric pressure is $8.59 \times 10^4 \text{ N/m}^2$? (b) What about at an altitude of 3000 m (about 10,000 ft) when atmospheric pressure is $7.00 \times 10^4 \text{ N/m}^2$?

Exercise:**Problem:**

What is the atmospheric pressure on top of Mt. Everest on a day when water boils there at a temperature of 70.0°C ?

Solution:

$$3.12 \times 10^4 \text{ Pa}$$

Exercise:**Problem:**

At a spot in the high Andes, water boils at 80.0°C , greatly reducing the cooking speed of potatoes, for example. What is atmospheric pressure at this location?

Exercise:**Problem:**

What is the relative humidity on a 25.0°C day when the air contains 18.0 g/m^3 of water vapor?

Solution:

$$78.3\%$$

Exercise:

Problem:

What is the density of water vapor in g/m^3 on a hot dry day in the desert when the temperature is 40.0°C and the relative humidity is 6.00%?

Exercise:**Problem:**

A deep-sea diver should breathe a gas mixture that has the same oxygen partial pressure as at sea level, where dry air contains 20.9% oxygen and has a total pressure of $1.01 \times 10^5 \text{ N/m}^2$. (a) What is the partial pressure of oxygen at sea level? (b) If the diver breathes a gas mixture at a pressure of $2.00 \times 10^6 \text{ N/m}^2$, what percent oxygen should it be to have the same oxygen partial pressure as at sea level?

Solution:

(a) $2.12 \times 10^4 \text{ Pa}$

(b) 1.06 %

Exercise:**Problem:**

The vapor pressure of water at 40.0°C is $7.34 \times 10^3 \text{ N/m}^2$. Using the ideal gas law, calculate the density of water vapor in g/m^3 that creates a partial pressure equal to this vapor pressure. The result should be the same as the saturation vapor density at that temperature (51.1 g/m^3).

Exercise:

Problem:

Air in human lungs has a temperature of 37.0°C and a saturation vapor density of 44.0 g/m^3 . (a) If 2.00 L of air is exhaled and very dry air inhaled, what is the maximum loss of water vapor by the person? (b) Calculate the partial pressure of water vapor having this density, and compare it with the vapor pressure of $6.31 \times 10^3\text{ N/m}^2$.

Solution:

(a) $8.80 \times 10^{-2}\text{ g}$

(b) $6.30 \times 10^3\text{ Pa}$; the two values are nearly identical.

Exercise:**Problem:**

If the relative humidity is 90.0% on a muggy summer morning when the temperature is 20.0°C , what will it be later in the day when the temperature is 30.0°C , assuming the water vapor density remains constant?

Exercise:**Problem:**

Late on an autumn day, the relative humidity is 45.0% and the temperature is 20.0°C . What will the relative humidity be that evening when the temperature has dropped to 10.0°C , assuming constant water vapor density?

Solution:

82.3%

Exercise:

Problem:

Atmospheric pressure atop Mt. Everest is $3.30 \times 10^4 \text{ N/m}^2$. (a) What is the partial pressure of oxygen there if it is 20.9% of the air? (b) What percent oxygen should a mountain climber breathe so that its partial pressure is the same as at sea level, where atmospheric pressure is $1.01 \times 10^5 \text{ N/m}^2$? (c) One of the most severe problems for those climbing very high mountains is the extreme drying of breathing passages. Why does this drying occur?

Exercise:**Problem:**

What is the dew point (the temperature at which 100% relative humidity would occur) on a day when relative humidity is 39.0% at a temperature of 20.0°C ?

Solution:

4.77°C

Exercise:**Problem:**

On a certain day, the temperature is 25.0°C and the relative humidity is 90.0%. How many grams of water must condense out of each cubic meter of air if the temperature falls to 15.0°C ? Such a drop in temperature can, thus, produce heavy dew or fog.

Exercise:**Problem: Integrated Concepts**

The boiling point of water increases with depth because pressure increases with depth. At what depth will fresh water have a boiling point of 150°C , if the surface of the water is at sea level?

Solution:

38.3 m

Exercise:**Problem: Integrated Concepts**

(a) At what depth in fresh water is the critical pressure of water reached, given that the surface is at sea level? (b) At what temperature will this water boil? (c) Is a significantly higher temperature needed to boil water at a greater depth?

Exercise:**Problem: Integrated Concepts**

To get an idea of the small effect that temperature has on Archimedes' principle, calculate the fraction of a copper block's weight that is supported by the buoyant force in 0°C water and compare this fraction with the fraction supported in 95.0°C water.

Solution:

$\frac{(F_B/w_{Cu})}{(F_B/w_{Cu})'} = 1.02$. The buoyant force supports nearly the exact same amount of force on the copper block in both circumstances.

Exercise:**Problem: Integrated Concepts**

If you want to cook in water at 150°C, you need a pressure cooker that can withstand the necessary pressure. (a) What pressure is required for the boiling point of water to be this high? (b) If the lid of the pressure cooker is a disk 25.0 cm in diameter, what force must it be able to withstand at this pressure?

Exercise:

Problem: Unreasonable Results

(a) How many moles per cubic meter of an ideal gas are there at a pressure of $1.00 \times 10^{14} \text{ N/m}^2$ and at 0°C ? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) $4.41 \times 10^{10} \text{ mol/m}^3$

(b) It's unreasonably large.

(c) At high pressures such as these, the ideal gas law can no longer be applied. As a result, unreasonable answers come up when it is used.

Exercise:**Problem: Unreasonable Results**

(a) An automobile mechanic claims that an aluminum rod fits loosely into its hole on an aluminum engine block because the engine is hot and the rod is cold. If the hole is 10.0% bigger in diameter than the 22.0°C rod, at what temperature will the rod be the same size as the hole? (b) What is unreasonable about this temperature? (c) Which premise is responsible?

Exercise:**Problem: Unreasonable Results**

The temperature inside a supernova explosion is said to be $2.00 \times 10^{13} \text{ K}$. (a) What would the average velocity v_{rms} of hydrogen atoms be? (b) What is unreasonable about this velocity? (c) Which premise or assumption is responsible?

Solution:

(a) $7.03 \times 10^8 \text{ m/s}$

(b) The velocity is too high—it's greater than the speed of light.

(c) The assumption that hydrogen inside a supernova behaves as an idea gas is responsible, because of the great temperature and density in the core of a star. Furthermore, when a velocity greater than the speed of light is obtained, classical physics must be replaced by relativity, a subject not yet covered.

Exercise:

Problem: Unreasonable Results

Suppose the relative humidity is 80% on a day when the temperature is 30.0°C . (a) What will the relative humidity be if the air cools to 25.0°C and the vapor density remains constant? (b) What is unreasonable about this result? (c) Which premise is responsible?

Glossary

dew point

the temperature at which relative humidity is 100%; the temperature at which water starts to condense out of the air

saturation

the condition of 100% relative humidity

percent relative humidity

the ratio of vapor density to saturation vapor density

relative humidity

the amount of water in the air relative to the maximum amount the air can hold

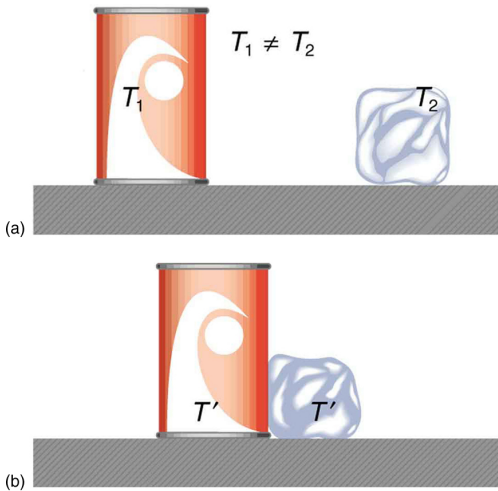
Heat

- Define heat as transfer of energy.

In [Work, Energy, and Energy Resources](#), we defined work as force times distance and learned that work done on an object changes its kinetic energy. We also saw in [Temperature, Kinetic Theory, and the Gas Laws](#) that temperature is proportional to the (average) kinetic energy of atoms and molecules. We say that a thermal system has a certain internal energy: its internal energy is higher if the temperature is higher. If two objects at different temperatures are brought in contact with each other, energy is transferred from the hotter to the colder object until equilibrium is reached and the bodies reach thermal equilibrium (i.e., they are at the same temperature). No work is done by either object, because no force acts through a distance. The transfer of energy is caused by the temperature difference, and ceases once the temperatures are equal. These observations lead to the following definition of **heat**: Heat is the spontaneous transfer of energy due to a temperature difference.

As noted in [Temperature, Kinetic Theory, and the Gas Laws](#), heat is often confused with temperature. For example, we may say the heat was unbearable, when we actually mean that the temperature was high. Heat is a form of energy, whereas temperature is not. The misconception arises because we are sensitive to the flow of heat, rather than the temperature.

Owing to the fact that heat is a form of energy, it has the SI unit of *joule* (J). The *calorie* (cal) is a common unit of energy, defined as the energy needed to change the temperature of 1.00 g of water by 1.00°C—specifically, between 14.5°C and 15.5°C, since there is a slight temperature dependence. Perhaps the most common unit of heat is the **kilocalorie** (kcal), which is the energy needed to change the temperature of 1.00 kg of water by 1.00°C. Since mass is most often specified in kilograms, kilocalorie is commonly used. Food calories (given the notation Cal, and sometimes called “big calorie”) are actually kilocalories (1 kilocalorie = 1000 calories), a fact not easily determined from package labeling.



In figure (a) the soft drink and the ice have different temperatures, T_1 and T_2 , and are not in thermal equilibrium. In figure (b), when the soft drink and ice are allowed to interact, energy is transferred until they reach the same temperature T' , achieving equilibrium. Heat transfer occurs due to the difference in temperatures. In fact, since the soft drink and ice are both in contact with the surrounding air and bench, the equilibrium temperature will be the same for both.

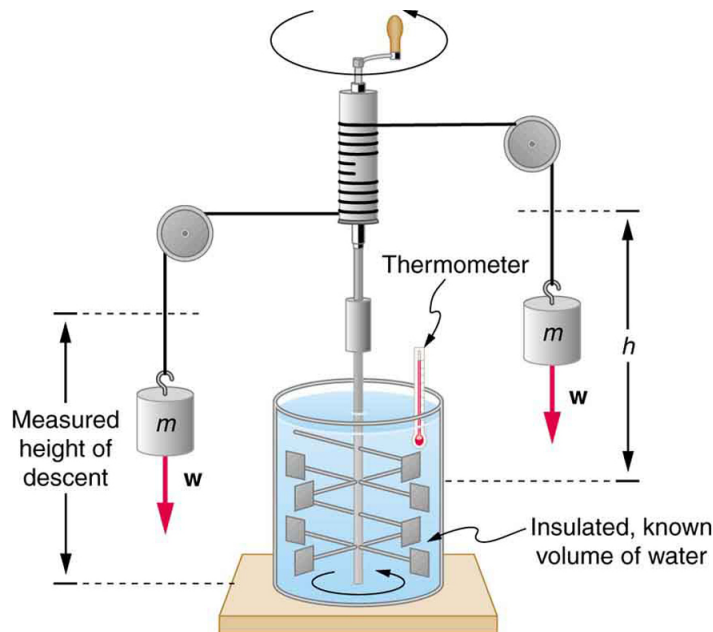
Mechanical Equivalent of Heat

It is also possible to change the temperature of a substance by doing work. Work can transfer energy into or out of a system. This realization helped establish the fact that heat is a form of energy. James Prescott Joule (1818–1889) performed many experiments to establish the **mechanical equivalent of heat**—*the work needed to produce the same effects as heat transfer*. In terms of the units used for these two terms, the best modern value for this equivalence is

Equation:

$$1.000 \text{ kcal} = 4186 \text{ J.}$$

We consider this equation as the conversion between two different units of energy.



Schematic depiction of Joule's experiment that established the equivalence of heat and work.

The figure above shows one of Joule's most famous experimental setups for demonstrating the mechanical equivalent of heat. It demonstrated that work and heat can produce the same effects, and helped establish the principle of conservation of energy. Gravitational potential energy (PE) (work done by the gravitational force) is converted into kinetic energy (KE), and then randomized by viscosity and turbulence into increased average kinetic energy of atoms and molecules in the system, producing a temperature increase. His contributions to the field of thermodynamics were so significant that the SI unit of energy was named after him.

Heat added or removed from a system changes its internal energy and thus its temperature. Such a temperature increase is observed while cooking. However, adding heat does not necessarily increase the temperature. An example is melting of ice; that is, when a substance changes from one phase to another. Work done on the system or by the system can also change the internal energy of the system. Joule demonstrated that the temperature of a system can be increased by stirring. If an ice cube is rubbed against a rough surface, work is done by the frictional force. A system has a well-defined internal energy, but we cannot say that it has a certain "heat content" or "work content". We use the phrase "heat transfer" to emphasize its nature.

Exercise:

Check Your Understanding

Problem:

Two samples (A and B) of the same substance are kept in a lab. Someone adds 10 kilojoules (kJ) of heat to one sample, while 10 kJ of work is done on the other sample. How can you tell to which sample the heat was added?

Solution:

Heat and work both change the internal energy of the substance. However, the properties of the sample only depend on the internal energy so that it is impossible to tell whether heat was added to sample A or B.

Summary

- Heat and work are the two distinct methods of energy transfer.
- Heat is energy transferred solely due to a temperature difference.
- Any energy unit can be used for heat transfer, and the most common are kilocalorie (kcal) and joule (J).
- Kilocalorie is defined to be the energy needed to change the temperature of 1.00 kg of water between 14.5°C and 15.5°C.
- The mechanical equivalent of this heat transfer is
 $1.00 \text{ kcal} = 4186 \text{ J}$.

Conceptual Questions

Exercise:

Problem: How is heat transfer related to temperature?

Exercise:

Problem:

Describe a situation in which heat transfer occurs. What are the resulting forms of energy?

Exercise:

Problem:

When heat transfers into a system, is the energy stored as heat? Explain briefly.

Glossary

heat

the spontaneous transfer of energy due to a temperature difference

kilocalorie

1 kilocalorie = 1000 calories

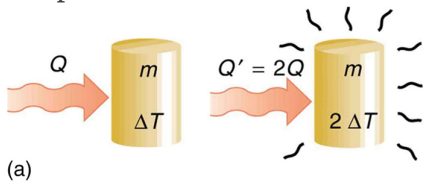
mechanical equivalent of heat

the work needed to produce the same effects as heat transfer

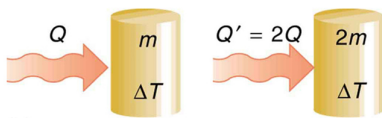
Temperature Change and Heat Capacity

- Observe heat transfer and change in temperature and mass.
- Calculate final temperature after heat transfer between two objects.

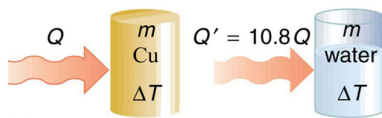
One of the major effects of heat transfer is temperature change: heating increases the temperature while cooling decreases it. We assume that there is no phase change and that no work is done on or by the system. Experiments show that the transferred heat depends on three factors—the change in temperature, the mass of the system, and the substance and phase of the substance.



(a)



(b)



(c)

The heat Q transferred to cause a temperature change depends on the magnitude of the temperature change, the mass of the system, and the substance and phase involved. (a) The amount of heat transferred is directly proportional to the temperature change.

To double the temperature change of a mass m , you need to add twice the heat. (b) The amount of heat transferred is also directly proportional to the mass.

To cause an equivalent temperature change in a

doubled mass, you need to add twice the heat. (c)

The amount of heat transferred depends on the substance and its phase. If it takes an amount Q of heat to cause a temperature change ΔT in a given mass of copper, it will take 10.8 times that amount of heat to cause the equivalent temperature change in the same mass of water assuming no phase change in either substance.

The dependence on temperature change and mass are easily understood. Owing to the fact that the (average) kinetic energy of an atom or molecule is proportional to the absolute temperature, the internal energy of a system is proportional to the absolute temperature and the number of atoms or molecules. Owing to the fact that the transferred heat is equal to the change in the internal energy, the heat is proportional to the mass of the substance and the temperature change. The transferred heat also depends on the substance so that, for example, the heat necessary to raise the temperature is less for alcohol than for water. For the same substance, the transferred heat also depends on the phase (gas, liquid, or solid).

Note:

Heat Transfer and Temperature Change

The quantitative relationship between heat transfer and temperature change contains all three factors:

Equation:

$$Q = mc\Delta T,$$

where Q is the symbol for heat transfer, m is the mass of the substance, and ΔT is the change in temperature. The symbol c stands for **specific heat** and depends on the material and phase. The specific heat is the amount of heat necessary to change the

temperature of 1.00 kg of mass by 1.00°C. The specific heat c is a property of the substance; its SI unit is J/(kg · K) or J/(kg · °C). Recall that the temperature change (ΔT) is the same in units of kelvin and degrees Celsius. If heat transfer is measured in kilocalories, then *the unit of specific heat* is kcal/(kg · °C).

Values of specific heat must generally be looked up in tables, because there is no simple way to calculate them. In general, the specific heat also depends on the temperature. [\[link\]](#) lists representative values of specific heat for various substances. Except for gases, the temperature and volume dependence of the specific heat of most substances is weak. We see from this table that the specific heat of water is five times that of glass and ten times that of iron, which means that it takes five times as much heat to raise the temperature of water the same amount as for glass and ten times as much heat to raise the temperature of water as for iron. In fact, water has one of the largest specific heats of any material, which is important for sustaining life on Earth.

Example:

Calculating the Required Heat: Heating Water in an Aluminum Pan

A 0.500 kg aluminum pan on a stove is used to heat 0.250 liters of water from 20.0°C to 80.0°C. (a) How much heat is required? What percentage of the heat is used to raise the temperature of (b) the pan and (c) the water?

Strategy

The pan and the water are always at the same temperature. When you put the pan on the stove, the temperature of the water and the pan is increased by the same amount. We use the equation for the heat transfer for the given temperature change and mass of water and aluminum. The specific heat values for water and aluminum are given in [\[link\]](#).

Solution

Because water is in thermal contact with the aluminum, the pan and the water are at the same temperature.

1. Calculate the temperature difference:

Equation:

$$\Delta T = T_f - T_i = 60.0^\circ\text{C}.$$

2. Calculate the mass of water. Because the density of water is 1000 kg/m³, one liter of water has a mass of 1 kg, and the mass of 0.250 liters of water is $m_w = 0.250$ kg.
3. Calculate the heat transferred to the water. Use the specific heat of water in [\[link\]](#):

Equation:

$$Q_w = m_w c_w \Delta T = (0.250 \text{ kg})(4186 \text{ J/kg}^\circ\text{C})(60.0^\circ\text{C}) = 62.8 \text{ kJ}.$$

4. Calculate the heat transferred to the aluminum. Use the specific heat for aluminum in [\[link\]](#):

Equation:

$$Q_{Al} = m_{Al} c_{Al} \Delta T = (0.500 \text{ kg})(900 \text{ J/kg}^\circ\text{C})(60.0^\circ\text{C}) = 27.0 \times 10^4 \text{ J} = 27.0 \text{ kJ}.$$

5. Compare the percentage of heat going into the pan versus that going into the water. First, find the total transferred heat:

Equation:

$$Q_{\text{Total}} = Q_W + Q_{Al} = 62.8 \text{ kJ} + 27.0 \text{ kJ} = 89.8 \text{ kJ}.$$

Thus, the amount of heat going into heating the pan is

Equation:

$$\frac{27.0 \text{ kJ}}{89.8 \text{ kJ}} \times 100\% = 30.1\%,$$

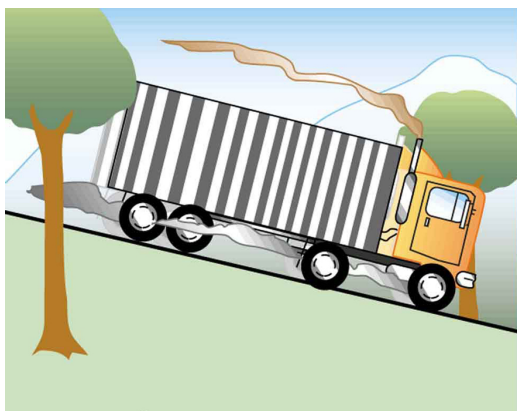
and the amount going into heating the water is

Equation:

$$\frac{62.8 \text{ kJ}}{89.8 \text{ kJ}} \times 100\% = 69.9\%.$$

Discussion

In this example, the heat transferred to the container is a significant fraction of the total transferred heat. Although the mass of the pan is twice that of the water, the specific heat of water is over four times greater than that of aluminum. Therefore, it takes a bit more than twice the heat to achieve the given temperature change for the water as compared to the aluminum pan.



The smoking brakes on this truck are a visible evidence of the mechanical equivalent of heat.

Example:

Calculating the Temperature Increase from the Work Done on a Substance: Truck Brakes Overheat on Downhill Runs

Truck brakes used to control speed on a downhill run do work, converting gravitational potential energy into increased internal energy (higher temperature) of the brake material. This conversion prevents the gravitational potential energy from being converted into kinetic energy of the truck. The problem is that the mass of the truck is large compared with that of the brake material absorbing the energy, and the temperature increase may occur too fast for sufficient heat to transfer from the brakes to the environment.

Calculate the temperature increase of 100 kg of brake material with an average specific heat of $800 \text{ J/kg} \cdot ^\circ\text{C}$ if the material retains 10% of the energy from a 10,000-kg truck descending 75.0 m (in vertical displacement) at a constant speed.

Strategy

If the brakes are not applied, gravitational potential energy is converted into kinetic energy. When brakes are applied, gravitational potential energy is converted into internal energy of the brake material. We first calculate the gravitational potential energy (Mgh) that the entire truck loses in its descent and then find the temperature increase produced in the brake material alone.

Solution

1. Calculate the change in gravitational potential energy as the truck goes downhill

Equation:

$$Mgh = (10,000 \text{ kg}) (9.80 \text{ m/s}^2) (75.0 \text{ m}) = 7.35 \times 10^6 \text{ J.}$$

2. Calculate the temperature from the heat transferred using $Q=Mgh$ and

Equation:

$$\Delta T = \frac{Q}{mc},$$

where m is the mass of the brake material. Insert the values $m = 100 \text{ kg}$ and $c = 800 \text{ J/kg} \cdot ^\circ\text{C}$ to find

Equation:

$$\Delta T = \frac{(7.35 \times 10^5 \text{ J})}{(100 \text{ kg})(800 \text{ J/kg}^\circ\text{C})} = 9.2^\circ\text{C.}$$

Discussion

This same idea underlies the recent hybrid technology of cars, where mechanical energy (gravitational potential energy) is converted by the brakes into electrical energy (battery).

Substances	Specific heat (c)	
Solids	J/kg·°C	kcal/kg·°C [footnote] These values are identical in units of cal/g ·°C.
Aluminum	900	0.215
Asbestos	800	0.19
Concrete, granite (average)	840	0.20
Copper	387	0.0924
Glass	840	0.20

Substances	Specific heat (c)	
Gold	129	0.0308
Human body (average at 37 °C)	3500	0.83
Ice (average, -50°C to 0°C)	2090	0.50
Iron, steel	452	0.108
Lead	128	0.0305
Silver	235	0.0562
Wood	1700	0.4
<i>Liquids</i>		
Benzene	1740	0.415
Ethanol	2450	0.586
Glycerin	2410	0.576
Mercury	139	0.0333
Water (15.0 °C)	4186	1.000
<i>Gases</i> [footnote] c_v at constant volume and at 20.0°C, except as noted, and at 1.00 atm average pressure. Values in parentheses are c_p at a constant pressure of 1.00 atm.		
Air (dry)	721 (1015)	0.172 (0.242)
Ammonia	1670 (2190)	0.399 (0.523)
Carbon dioxide	638 (833)	0.152 (0.199)

Substances	Specific heat (c)	
Nitrogen	739 (1040)	0.177 (0.248)
Oxygen	651 (913)	0.156 (0.218)
Steam (100°C)	1520 (2020)	0.363 (0.482)

Specific Heats^[footnote] of Various Substances

The values for solids and liquids are at constant volume and at 25°C, except as noted.

Note that [\[link\]](#) is an illustration of the mechanical equivalent of heat. Alternatively, the temperature increase could be produced by a blow torch instead of mechanically.

Example:

Calculating the Final Temperature When Heat Is Transferred Between Two Bodies: Pouring Cold Water in a Hot Pan

Suppose you pour 0.250 kg of 20.0°C water (about a cup) into a 0.500-kg aluminum pan off the stove with a temperature of 150°C. Assume that the pan is placed on an insulated pad and that a negligible amount of water boils off. What is the temperature when the water and pan reach thermal equilibrium a short time later?

Strategy

The pan is placed on an insulated pad so that little heat transfer occurs with the surroundings. Originally the pan and water are not in thermal equilibrium: the pan is at a higher temperature than the water. Heat transfer then restores thermal equilibrium once the water and pan are in contact. Because heat transfer between the pan and water takes place rapidly, the mass of evaporated water is negligible and the magnitude of the heat lost by the pan is equal to the heat gained by the water. The exchange of heat stops once a thermal equilibrium between the pan and the water is achieved. The heat exchange can be written as $|Q_{\text{hot}}| = Q_{\text{cold}}$.

Solution

1. Use the equation for heat transfer $Q = mc\Delta T$ to express the heat lost by the aluminum pan in terms of the mass of the pan, the specific heat of aluminum, the initial temperature of the pan, and the final temperature:

Equation:

$$Q_{\text{hot}} = m_{\text{Al}}c_{\text{Al}}(T_{\text{f}} - 150^{\circ}\text{C}).$$

2. Express the heat gained by the water in terms of the mass of the water, the specific heat of water, the initial temperature of the water and the final temperature:

Equation:

$$Q_{\text{cold}} = m_{\text{W}}c_{\text{W}}(T_{\text{f}} - 20.0^{\circ}\text{C}).$$

3. Note that $Q_{\text{hot}} < 0$ and $Q_{\text{cold}} > 0$ and that they must sum to zero because the heat lost by the hot pan must be the same as the heat gained by the cold water:

Equation:

$$\begin{aligned}Q_{\text{cold}} + Q_{\text{hot}} &= 0, \\Q_{\text{cold}} &= -Q_{\text{hot}}, \\m_{\text{W}}c_{\text{W}}(T_{\text{f}} - 20.0^{\circ}\text{C}) &= -m_{\text{Al}}c_{\text{Al}}(T_{\text{f}} - 150^{\circ}\text{C}).\end{aligned}$$

4. Bring all terms involving T_{f} on the left hand side and all other terms on the right hand side. Solve for T_{f} ,

Equation:

$$T_{\text{f}} = \frac{m_{\text{Al}}c_{\text{Al}}(150^{\circ}\text{C}) + m_{\text{W}}c_{\text{W}}(20.0^{\circ}\text{C})}{m_{\text{Al}}c_{\text{Al}} + m_{\text{W}}c_{\text{W}}},$$

and insert the numerical values:

Equation:

$$\begin{aligned}T_{\text{f}} &= \frac{(0.500 \text{ kg})(900 \text{ J/kg}^{\circ}\text{C})(150^{\circ}\text{C}) + (0.250 \text{ kg})(4186 \text{ J/kg}^{\circ}\text{C})(20.0^{\circ}\text{C})}{(0.500 \text{ kg})(900 \text{ J/kg}^{\circ}\text{C}) + (0.250 \text{ kg})(4186 \text{ J/kg}^{\circ}\text{C})} \\&= \frac{88430 \text{ J}}{1496.5 \text{ J/}^{\circ}\text{C}} \\&= 59.1^{\circ}\text{C}.\end{aligned}$$

Discussion

This is a typical *calorimetry* problem—two bodies at different temperatures are brought in contact with each other and exchange heat until a common temperature is reached. Why is the final temperature so much closer to 20.0°C than 150°C ? The reason is that water has a greater specific heat than most common substances and thus undergoes a small temperature change for a given heat transfer. A large body of water, such as a lake, requires a large amount of heat to increase its temperature appreciably. This explains why the temperature of a lake stays relatively constant during a day even when the temperature change of the air is large. However, the water temperature does change over longer times (e.g., summer to winter).

Note:**Take-Home Experiment: Temperature Change of Land and Water**

What heats faster, land or water?

To study differences in heat capacity:

- Place equal masses of dry sand (or soil) and water at the same temperature into two small jars. (The average density of soil or sand is about 1.6 times that of water, so you can achieve approximately equal masses by using 50% more water by volume.)
- Heat both (using an oven or a heat lamp) for the same amount of time.
- Record the final temperature of the two masses.
- Now bring both jars to the same temperature by heating for a longer period of time.
- Remove the jars from the heat source and measure their temperature every 5 minutes for about 30 minutes.

Which sample cools off the fastest? This activity replicates the phenomena responsible for land breezes and sea breezes.

Exercise:**Check Your Understanding****Problem:**

If 25 kJ is necessary to raise the temperature of a block from 25°C to 30°C, how much heat is necessary to heat the block from 45°C to 50°C?

Solution:

The heat transfer depends only on the temperature difference. Since the temperature differences are the same in both cases, the same 25 kJ is necessary in the second case.

Summary

- The transfer of heat Q that leads to a change ΔT in the temperature of a body with mass m is $Q = mc\Delta T$, where c is the specific heat of the material. This relationship can also be considered as the definition of specific heat.

Conceptual Questions**Exercise:**

Problem:

What three factors affect the heat transfer that is necessary to change an object's temperature?

Exercise:**Problem:**

The brakes in a car increase in temperature by ΔT when bringing the car to rest from a speed v . How much greater would ΔT be if the car initially had twice the speed? You may assume the car to stop sufficiently fast so that no heat transfers out of the brakes.

Problems & Exercises**Exercise:****Problem:**

On a hot day, the temperature of an 80,000-L swimming pool increases by 1.50°C . What is the net heat transfer during this heating? Ignore any complications, such as loss of water by evaporation.

Solution:**Equation:**

$$5.02 \times 10^8 \text{ J}$$

Exercise:

Problem: Show that $1 \text{ cal/g} \cdot ^\circ\text{C} = 1 \text{ kcal/kg} \cdot ^\circ\text{C}$.

Exercise:**Problem:**

To sterilize a 50.0-g glass baby bottle, we must raise its temperature from 22.0°C to 95.0°C . How much heat transfer is required?

Solution:**Equation:**

$$3.07 \times 10^3 \text{ J}$$

Exercise:**Problem:**

The same heat transfer into identical masses of different substances produces different temperature changes. Calculate the final temperature when 1.00 kcal of heat transfers into 1.00 kg of the following, originally at 20.0°C: (a) water; (b) concrete; (c) steel; and (d) mercury.

Exercise:**Problem:**

Rubbing your hands together warms them by converting work into thermal energy. If a woman rubs her hands back and forth for a total of 20 rubs, at a distance of 7.50 cm per rub, and with an average frictional force of 40.0 N, what is the temperature increase? The mass of tissues warmed is only 0.100 kg, mostly in the palms and fingers.

Solution:**Equation:**

$$0.171^{\circ}\text{C}$$

Exercise:**Problem:**

A 0.250-kg block of a pure material is heated from 20.0°C to 65.0°C by the addition of 4.35 kJ of energy. Calculate its specific heat and identify the substance of which it is most likely composed.

Exercise:**Problem:**

Suppose identical amounts of heat transfer into different masses of copper and water, causing identical changes in temperature. What is the ratio of the mass of copper to water?

Solution:

$$10.8$$

Exercise:

Problem:

(a) The number of kilocalories in food is determined by calorimetry techniques in which the food is burned and the amount of heat transfer is measured. How many kilocalories per gram are there in a 5.00-g peanut if the energy from burning it is transferred to 0.500 kg of water held in a 0.100-kg aluminum cup, causing a 54.9°C temperature increase? (b) Compare your answer to labeling information found on a package of peanuts and comment on whether the values are consistent.

Exercise:**Problem:**

Following vigorous exercise, the body temperature of an 80.0-kg person is 40.0°C . At what rate in watts must the person transfer thermal energy to reduce the body temperature to 37.0°C in 30.0 min, assuming the body continues to produce energy at the rate of 150 W? (1 watt = 1 joule/second or $1\text{ W} = 1\text{ J/s}$).

Solution:

617 W

Exercise:**Problem:**

Even when shut down after a period of normal use, a large commercial nuclear reactor transfers thermal energy at the rate of 150 MW by the radioactive decay of fission products. This heat transfer causes a rapid increase in temperature if the cooling system fails

(1 watt = 1 joule/second or $1\text{ W} = 1\text{ J/s}$ and $1\text{ MW} = 1\text{ megawatt}$). (a) Calculate the rate of temperature increase in degrees Celsius per second ($^{\circ}\text{C/s}$) if the mass of the reactor core is $1.60 \times 10^5\text{ kg}$ and it has an average specific heat of $0.3349\text{ kJ/kg}^{\circ}\text{C}$. (b) How long would it take to obtain a temperature increase of 2000°C , which could cause some metals holding the radioactive materials to melt? (The initial rate of temperature increase would be greater than that calculated here because the heat transfer is concentrated in a smaller mass. Later, however, the temperature increase would slow down because the $5 \times 10^5\text{-kg}$ steel containment vessel would also begin to heat up.)



Radioactive spent-fuel pool at a nuclear power plant. Spent fuel stays hot for a long time. (credit: U.S. Department of Energy)

Glossary

specific heat

the amount of heat necessary to change the temperature of 1.00 kg of a substance by 1.00 °C

Phase Change and Latent Heat

- Examine heat transfer.
- Calculate final temperature from heat transfer.

So far we have discussed temperature change due to heat transfer. No temperature change occurs from heat transfer if ice melts and becomes liquid water (i.e., during a phase change). For example, consider water dripping from icicles melting on a roof warmed by the Sun. Conversely, water freezes in an ice tray cooled by lower-temperature surroundings.



Heat from the air transfers to the ice causing it to melt.
(credit: Mike Brand)

Energy is required to melt a solid because the cohesive bonds between the molecules in the solid must be broken apart such that, in the liquid, the molecules can move around at comparable kinetic energies; thus, there is no rise in temperature. Similarly, energy is needed to vaporize a liquid, because molecules in a liquid interact with each other via attractive forces. There is no temperature change until a phase change is complete. The temperature of a cup of soda initially at 0°C stays at 0°C until all the ice has melted. Conversely, energy is released during freezing and condensation, usually in the form of thermal energy. Work is done by cohesive forces when molecules are brought together. The corresponding energy must be given off (dissipated) to allow them to stay together [\[link\]](#).

The energy involved in a phase change depends on two major factors: the number and strength of bonds or force pairs. The number of bonds is proportional to the number of molecules and thus to the mass of the sample. The strength of forces depends on the type of molecules. The heat Q required to change the phase of a sample of mass m is given by

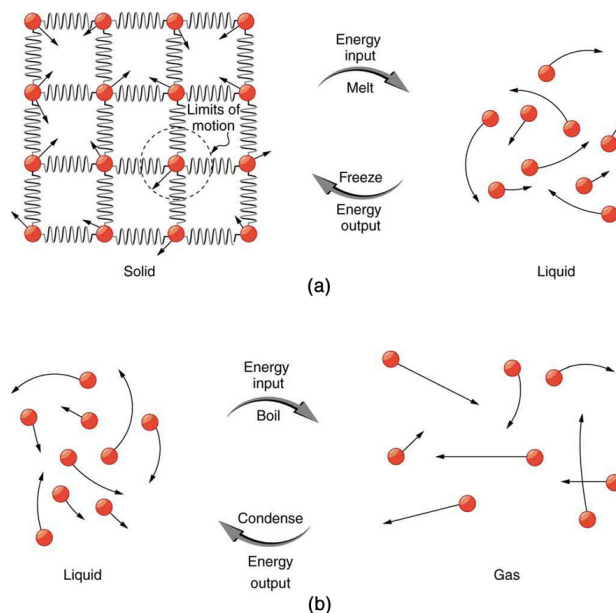
Equation:

$$Q = mL_f \text{ (melting/freezing),}$$

Equation:

$$Q = mL_v \text{ (vaporization/condensation),}$$

where the latent heat of fusion, L_f , and latent heat of vaporization, L_v , are material constants that are determined experimentally. See ([link](#)).



(a) Energy is required to partially overcome the attractive forces between molecules in a solid to form a liquid. That same energy must be removed for freezing to take place. (b) Molecules are separated by large distances when going from liquid to vapor, requiring significant energy to overcome molecular attraction. The same energy must be removed for condensation to take place. There is no temperature change until a phase change is complete.

Latent heat is measured in units of J/kg. Both L_f and L_v depend on the substance, particularly on the strength of its molecular forces as noted earlier. L_f and L_v are collectively called **latent heat coefficients**. They are *latent*, or hidden, because in phase changes, energy enters or leaves a system without causing a temperature change in the system; so, in effect, the energy is hidden. [\[link\]](#) lists representative values of L_f and L_v , together with melting and boiling points.

The table shows that significant amounts of energy are involved in phase changes. Let us look, for example, at how much energy is needed to melt a kilogram of ice at 0°C to produce a kilogram of water at 0°C . Using the equation for a change in temperature and the value for water from [\[link\]](#), we find that $Q = mL_f = (1.0\text{ kg})(334\text{ kJ/kg}) = 334\text{ kJ}$ is the energy to melt a kilogram of ice. This is a lot of energy as it represents the same amount of energy needed to raise the temperature of 1 kg of liquid water from 0°C to 79.8°C . Even more energy is required to vaporize water; it would take 2256 kJ to change 1 kg of liquid water at the normal boiling point (100°C at atmospheric pressure) to steam (water vapor). This example shows that the energy for a phase change is enormous compared to energy associated with temperature changes without a phase change.

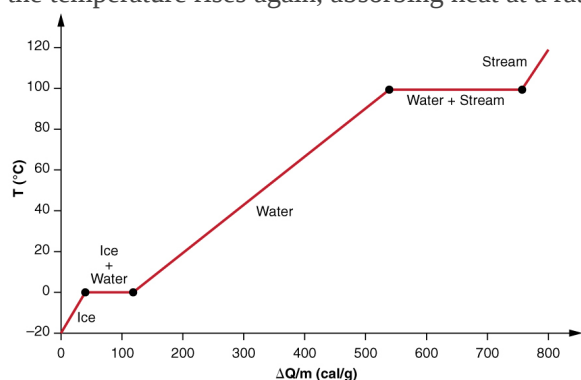
		L_f			L_v	
Substance	Melting point (°C)	kJ/kg	kcal/kg	Boiling point (°C)	kJ/kg	kcal/kg
Helium	−269.7	5.23	1.25	−268.9	20.9	4.99
Hydrogen	−259.3	58.6	14.0	−252.9	452	108
Nitrogen	−210.0	25.5	6.09	−195.8	201	48.0
Oxygen	−218.8	13.8	3.30	−183.0	213	50.9
Ethanol	−114	104	24.9	78.3	854	204
Ammonia	−75		108	−33.4	1370	327
Mercury	−38.9	11.8	2.82	357	272	65.0
Water	0.00	334	79.8	100.0	2256 ^[footnote] At 37.0°C (body temperature), the heat of vaporization L_v for water is 2430 kJ/kg or 580 kcal/kg	539 ^[footnote] At 37.0°C (body temperature), the heat of vaporization L_v for water is 2430 kJ/kg or 580 kcal/kg
Sulfur	119	38.1	9.10	444.6	326	77.9
Lead	327	24.5	5.85	1750	871	208
Antimony	631	165	39.4	1440	561	134
Aluminum	660	380	90	2450	11400	2720
Silver	961	88.3	21.1	2193	2336	558
Gold	1063	64.5	15.4	2660	1578	377
Copper	1083	134	32.0	2595	5069	1211
Uranium	1133	84	20	3900	1900	454
Tungsten	3410	184	44	5900	4810	1150

Heats of Fusion and Vaporization ^[footnote]

Values quoted at the normal melting and boiling temperatures at standard atmospheric pressure (1 atm).

Phase changes can have a tremendous stabilizing effect even on temperatures that are not near the melting and boiling points, because evaporation and condensation (conversion of a gas into a liquid state) occur even at temperatures below the boiling point. Take, for example, the fact that air temperatures in humid climates rarely go above 35.0°C , which is because most heat transfer goes into evaporating water into the air. Similarly, temperatures in humid weather rarely fall below the dew point because enormous heat is released when water vapor condenses.

We examine the effects of phase change more precisely by considering adding heat into a sample of ice at -20°C ([link](#)). The temperature of the ice rises linearly, absorbing heat at a constant rate of $0.50\text{ cal/g} \cdot ^{\circ}\text{C}$ until it reaches 0°C . Once at this temperature, the ice begins to melt until all the ice has melted, absorbing 79.8 cal/g of heat. The temperature remains constant at 0°C during this phase change. Once all the ice has melted, the temperature of the liquid water rises, absorbing heat at a new constant rate of $1.00\text{ cal/g} \cdot ^{\circ}\text{C}$. At 100°C , the water begins to boil and the temperature again remains constant while the water absorbs 539 cal/g of heat during this phase change. When all the liquid has become steam vapor, the temperature rises again, absorbing heat at a rate of $0.482\text{ cal/g} \cdot ^{\circ}\text{C}$.



A graph of temperature versus energy added. The system is constructed so that no vapor evaporates while ice warms to become liquid water, and so that, when vaporization occurs, the vapor remains in of the system. The long stretches of constant temperature values at 0°C and 100°C reflect the large latent heat of melting and vaporization, respectively.

Water can evaporate at temperatures below the boiling point. More energy is required than at the boiling point, because the kinetic energy of water molecules at temperatures below 100°C is less than that at 100°C , hence less energy is available from random thermal motions. Take, for example, the fact that, at body temperature, perspiration from the skin requires a heat input of 2428 kJ/kg , which is about 10 percent higher than the latent heat of vaporization at 100°C . This heat comes from the skin, and thus provides an effective cooling mechanism in hot weather. High humidity inhibits evaporation, so that body temperature might rise, leaving unevaporated sweat on your brow.

Example:
Calculate Final Temperature from Phase Change: Cooling Soda with Ice Cubes

Three ice cubes are used to chill a soda at 20°C with mass $m_{\text{soda}} = 0.25 \text{ kg}$. The ice is at 0°C and each ice cube has a mass of 6.0 g. Assume that the soda is kept in a foam container so that heat loss can be ignored. Assume the soda has the same heat capacity as water. Find the final temperature when all ice has melted.

Strategy

The ice cubes are at the melting temperature of 0°C. Heat is transferred from the soda to the ice for melting. Melting of ice occurs in two steps: first the phase change occurs and solid (ice) transforms into liquid water at the melting temperature, then the temperature of this water rises. Melting yields water at 0°C, so more heat is transferred from the soda to this water until the water plus soda system reaches thermal equilibrium,

Equation:

$$Q_{\text{ice}} = -Q_{\text{soda}}.$$

The heat transferred to the ice is $Q_{\text{ice}} = m_{\text{ice}}L_f + m_{\text{ice}}c_W(T_f - 0^\circ\text{C})$. The heat given off by the soda is $Q_{\text{soda}} = m_{\text{soda}}c_W(T_f - 20^\circ\text{C})$. Since no heat is lost, $Q_{\text{ice}} = -Q_{\text{soda}}$, so that

Equation:

$$m_{\text{ice}}L_f + m_{\text{ice}}c_W(T_f - 0^\circ\text{C}) = -m_{\text{soda}}c_W(T_f - 20^\circ\text{C}).$$

Bring all terms involving T_f on the left-hand-side and all other terms on the right-hand-side. Solve for the unknown quantity T_f :

Equation:

$$T_f = \frac{m_{\text{soda}}c_W(20^\circ\text{C}) - m_{\text{ice}}L_f}{(m_{\text{soda}} + m_{\text{ice}})c_W}.$$

Solution

1. Identify the known quantities. The mass of ice is $m_{\text{ice}} = 3 \times 6.0 \text{ g} = 0.018 \text{ kg}$ and the mass of soda is $m_{\text{soda}} = 0.25 \text{ kg}$.

2. Calculate the terms in the numerator:

Equation:

$$m_{\text{soda}}c_W(20^\circ\text{C}) = (0.25 \text{ kg})(4186 \text{ J/kg} \cdot ^\circ\text{C})(20^\circ\text{C}) = 20,930 \text{ J}$$

and

Equation:

$$m_{\text{ice}}L_f = (0.018 \text{ kg})(334,000 \text{ J/kg}) = 6012 \text{ J}.$$

3. Calculate the denominator:

Equation:

$$(m_{\text{soda}} + m_{\text{ice}})c_W = (0.25 \text{ kg} + 0.018 \text{ kg})(4186 \text{ J/(kg} \cdot ^\circ\text{C)}) = 1122 \text{ J/}^\circ\text{C}.$$

4. Calculate the final temperature:

Equation:

$$T_f = \frac{20,930 \text{ J} - 6012 \text{ J}}{1122 \text{ J/}^\circ\text{C}} = 13^\circ\text{C}.$$

Discussion

This example illustrates the enormous energies involved during a phase change. The mass of ice is about 7 percent the mass of water but leads to a noticeable change in the temperature of soda. Although we assumed that the ice was at the freezing temperature, this is incorrect: the typical temperature is -6°C . However, this correction gives a final temperature that is essentially identical to the result we found. Can you explain why?

We have seen that vaporization requires heat transfer to a liquid from the surroundings, so that energy is released by the surroundings. Condensation is the reverse process, increasing the temperature of the surroundings. This increase may seem surprising, since we associate condensation with cold objects—the glass in the figure, for example. However, energy must be removed from the condensing molecules to make a vapor condense. The energy is exactly the same as that required to make the phase change in the other direction, from liquid to vapor, and so it can be calculated from $Q = mL_v$.



Condensation forms on this glass of iced tea because the temperature of the nearby air is reduced to below the dew point. The rate at which water molecules join together exceeds the rate at which they separate, and so water condenses. Energy is released when the water condenses, speeding the melting of the ice in the glass. (credit: Jenny Downing)

Note:

Real-World Application

Energy is also released when a liquid freezes. This phenomenon is used by fruit growers in Florida to protect oranges when the temperature is close to the freezing point (0°C). Growers spray water on the

plants in orchards so that the water freezes and heat is released to the growing oranges on the trees. This prevents the temperature inside the orange from dropping below freezing, which would damage the fruit.



The ice on these trees released large amounts of energy when it froze, helping to prevent the temperature of the trees from dropping below 0°C .

Water is intentionally sprayed on orchards to help prevent hard frosts.

(credit: Hermann Hammer)

Sublimation is the transition from solid to vapor phase. You may have noticed that snow can disappear into thin air without a trace of liquid water, or the disappearance of ice cubes in a freezer. The reverse is also true: Frost can form on very cold windows without going through the liquid stage. A popular effect is the making of “smoke” from dry ice, which is solid carbon dioxide. Sublimation occurs because the equilibrium vapor pressure of solids is not zero. Certain air fresheners use the sublimation of a solid to inject a perfume into the room. Moth balls are a slightly toxic example of a phenol (an organic compound) that sublimates, while some solids, such as osmium tetroxide, are so toxic that they must be kept in sealed containers to prevent human exposure to their sublimation-produced vapors.



(a)



(b)

Direct transitions
between solid and

vapor are common,
sometimes useful,
and even beautiful.

(a) Dry ice
sublimates directly to
carbon dioxide gas.
The visible vapor is
made of water
droplets. (credit:
Windell Oskay) (b)
Frost forms patterns
on a very cold
window, an example
of a solid formed
directly from a
vapor. (credit: Liz
West)

All phase transitions involve heat. In the case of direct solid-vapor transitions, the energy required is given by the equation $Q = mL_s$, where L_s is the **heat of sublimation**, which is the energy required to change 1.00 kg of a substance from the solid phase to the vapor phase. L_s is analogous to L_f and L_v , and its value depends on the substance. Sublimation requires energy input, so that dry ice is an effective coolant, whereas the reverse process (i.e., frosting) releases energy. The amount of energy required for sublimation is of the same order of magnitude as that for other phase transitions.

The material presented in this section and the preceding section allows us to calculate any number of effects related to temperature and phase change. In each case, it is necessary to identify which temperature and phase changes are taking place and then to apply the appropriate equation. Keep in mind that heat transfer and work can cause both temperature and phase changes.

Problem-Solving Strategies for the Effects of Heat Transfer

1. *Examine the situation to determine that there is a change in the temperature or phase. Is there heat transfer into or out of the system?* When the presence or absence of a phase change is not obvious, you may wish to first solve the problem as if there were no phase changes, and examine the temperature change obtained. If it is sufficient to take you past a boiling or melting point, you should then go back and do the problem in steps—temperature change, phase change, subsequent temperature change, and so on.
2. *Identify and list all objects that change temperature and phase.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns).* A written list is useful.
4. *Make a list of what is given or what can be inferred from the problem as stated (identify the knowns).*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).* If there is a temperature change, the transferred heat depends on the specific heat (see [\[link\]](#)) whereas, for a phase change, the transferred heat depends on the latent heat. See [\[link\]](#).
6. *Substitute the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units.* You will need to do this in steps if there is more than one stage to the process (such as a temperature change followed by a phase change).

7. Check the answer to see if it is reasonable: Does it make sense? As an example, be certain that the temperature change does not also cause a phase change that you have not taken into account.

Exercise:

Check Your Understanding

Problem:

Why does snow remain on mountain slopes even when daytime temperatures are higher than the freezing temperature?

Solution:

Snow is formed from ice crystals and thus is the solid phase of water. Because enormous heat is necessary for phase changes, it takes a certain amount of time for this heat to be accumulated from the air, even if the air is above 0°C. The warmer the air is, the faster this heat exchange occurs and the faster the snow melts.

Summary

- Most substances can exist either in solid, liquid, and gas forms, which are referred to as “phases.”
- Phase changes occur at fixed temperatures for a given substance at a given pressure, and these temperatures are called boiling and freezing (or melting) points.
- During phase changes, heat absorbed or released is given by:

Equation:

$$Q = mL,$$

where L is the latent heat coefficient.

Conceptual Questions

Exercise:

Problem:

Heat transfer can cause temperature and phase changes. What else can cause these changes?

Exercise:

Problem:

How does the latent heat of fusion of water help slow the decrease of air temperatures, perhaps preventing temperatures from falling significantly below 0°C, in the vicinity of large bodies of water?

Exercise:

Problem: What is the temperature of ice right after it is formed by freezing water?

Exercise:

Problem:

If you place 0°C ice into 0°C water in an insulated container, what will happen? Will some ice melt, will more water freeze, or will neither take place?

Exercise:

Problem:

What effect does condensation on a glass of ice water have on the rate at which the ice melts? Will the condensation speed up the melting process or slow it down?

Exercise:

Problem:

In very humid climates where there are numerous bodies of water, such as in Florida, it is unusual for temperatures to rise above about 35°C (95°F). In deserts, however, temperatures can rise far above this. Explain how the evaporation of water helps limit high temperatures in humid climates.

Exercise:

Problem:

In winters, it is often warmer in San Francisco than in nearby Sacramento, 150 km inland. In summers, it is nearly always hotter in Sacramento. Explain how the bodies of water surrounding San Francisco moderate its extreme temperatures.

Exercise:

Problem:

Putting a lid on a boiling pot greatly reduces the heat transfer necessary to keep it boiling. Explain why.

Exercise:

Problem:

Freeze-dried foods have been dehydrated in a vacuum. During the process, the food freezes and must be heated to facilitate dehydration. Explain both how the vacuum speeds up dehydration and why the food freezes as a result.

Exercise:

Problem:

When still air cools by radiating at night, it is unusual for temperatures to fall below the dew point. Explain why.

Exercise:

Problem:

In a physics classroom demonstration, an instructor inflates a balloon by mouth and then cools it in liquid nitrogen. When cold, the shrunken balloon has a small amount of light blue liquid in it, as well as some snow-like crystals. As it warms up, the liquid boils, and part of the crystals sublime, with some crystals lingering for awhile and then producing a liquid. Identify the blue liquid and the two solids in the cold balloon. Justify your identifications using data from [\[link\]](#).

Problems & Exercises

Exercise:

Problem:

How much heat transfer (in kilocalories) is required to thaw a 0.450-kg package of frozen vegetables originally at 0°C if their heat of fusion is the same as that of water?

Solution:

35.9 kcal

Exercise:

Problem:

A bag containing 0°C ice is much more effective in absorbing energy than one containing the same amount of 0°C water.

- How much heat transfer is necessary to raise the temperature of 0.800 kg of water from 0°C to 30.0°C ?
- How much heat transfer is required to first melt 0.800 kg of 0°C ice and then raise its temperature?
- Explain how your answer supports the contention that the ice is more effective.

Exercise:

Problem:

- How much heat transfer is required to raise the temperature of a 0.750-kg aluminum pot containing 2.50 kg of water from 30.0°C to the boiling point and then boil away 0.750 kg of water?
 - How long does this take if the rate of heat transfer is 500 W
 $1 \text{ watt} = 1 \text{ joule/second}$ ($1 \text{ W} = 1 \text{ J/s}$)?
-

Solution:

(a) 591 kcal

(b) $4.94 \times 10^3 \text{ s}$

Exercise:

Problem:

The formation of condensation on a glass of ice water causes the ice to melt faster than it would otherwise. If 8.00 g of condensation forms on a glass containing both water and 200 g of ice, how many grams of the ice will melt as a result? Assume no other heat transfer occurs.

Exercise:

Problem:

On a trip, you notice that a 3.50-kg bag of ice lasts an average of one day in your cooler. What is the average power in watts entering the ice if it starts at 0°C and completely melts to 0°C water in exactly one day $1 \text{ watt} = 1 \text{ joule/second}$ ($1 \text{ W} = 1 \text{ J/s}$)?

Solution:

13.5 W

Exercise:**Problem:**

On a certain dry sunny day, a swimming pool's temperature would rise by 1.50°C if not for evaporation. What fraction of the water must evaporate to carry away precisely enough energy to keep the temperature constant?

Exercise:**Problem:**

- (a) How much heat transfer is necessary to raise the temperature of a 0.200-kg piece of ice from -20.0°C to 130°C , including the energy needed for phase changes?
- (b) How much time is required for each stage, assuming a constant 20.0 kJ/s rate of heat transfer?
- (c) Make a graph of temperature versus time for this process.

Solution:

(a) 148 kcal

(b) 0.418 s, 3.34 s, 4.19 s, 22.6 s, 0.456 s

Exercise:**Problem:**

In 1986, a gargantuan iceberg broke away from the Ross Ice Shelf in Antarctica. It was approximately a rectangle 160 km long, 40.0 km wide, and 250 m thick.

- (a) What is the mass of this iceberg, given that the density of ice is 917 kg/m^3 ?
- (b) How much heat transfer (in joules) is needed to melt it?
- (c) How many years would it take sunlight alone to melt ice this thick, if the ice absorbs an average of 100 W/m^2 , 12.00 h per day?

Exercise:**Problem:**

How many grams of coffee must evaporate from 350 g of coffee in a 100-g glass cup to cool the coffee from 95.0°C to 45.0°C ? You may assume the coffee has the same thermal properties as water and that the average heat of vaporization is 2340 kJ/kg (560 cal/g). (You may neglect the change in mass of the coffee as it cools, which will give you an answer that is slightly larger than correct.)

Solution:

33.0 g

Exercise:

Problem:

- (a) It is difficult to extinguish a fire on a crude oil tanker, because each liter of crude oil releases 2.80×10^7 J of energy when burned. To illustrate this difficulty, calculate the number of liters of water that must be expended to absorb the energy released by burning 1.00 L of crude oil, if the water has its temperature raised from 20.0°C to 100°C , it boils, and the resulting steam is raised to 300°C .
- (b) Discuss additional complications caused by the fact that crude oil has a smaller density than water.
-

Solution:

(a) 9.67 L

(b) Crude oil is less dense than water, so it floats on top of the water, thereby exposing it to the oxygen in the air, which it uses to burn. Also, if the water is under the oil, it is less efficient in absorbing the heat generated by the oil.

Exercise:**Problem:**

The energy released from condensation in thunderstorms can be very large. Calculate the energy released into the atmosphere for a small storm of radius 1 km, assuming that 1.0 cm of rain is precipitated uniformly over this area.

Exercise:

Problem: To help prevent frost damage, 4.00 kg of 0°C water is sprayed onto a fruit tree.

- (a) How much heat transfer occurs as the water freezes?
- (b) How much would the temperature of the 200-kg tree decrease if this amount of heat transferred from the tree? Take the specific heat to be $3.35 \text{ kJ/kg} \cdot ^\circ\text{C}$, and assume that no phase change occurs.
-

Solution:

a) 319 kcal

b) 2.00°C

Exercise:**Problem:**

A 0.250-kg aluminum bowl holding 0.800 kg of soup at 25.0°C is placed in a freezer. What is the final temperature if 377 kJ of energy is transferred from the bowl and soup, assuming the soup's thermal properties are the same as that of water? Explicitly show how you follow the steps in [Problem-Solving Strategies for the Effects of Heat Transfer](#).

Exercise:**Problem:**

A 0.0500-kg ice cube at -30.0°C is placed in 0.400 kg of 35.0°C water in a very well-insulated container. What is the final temperature?

Solution:

20.6°C

Exercise:**Problem:**

If you pour 0.0100 kg of 20.0°C water onto a 1.20-kg block of ice (which is initially at -15.0°C), what is the final temperature? You may assume that the water cools so rapidly that effects of the surroundings are negligible.

Exercise:**Problem:**

Indigenous people sometimes cook in watertight baskets by placing hot rocks into water to bring it to a boil. What mass of 500°C rock must be placed in 4.00 kg of 15.0°C water to bring its temperature to 100°C, if 0.0250 kg of water escapes as vapor from the initial sizzle? You may neglect the effects of the surroundings and take the average specific heat of the rocks to be that of granite.

Solution:

4.38 kg

Exercise:**Problem:**

What would be the final temperature of the pan and water in [Calculating the Final Temperature When Heat Is Transferred Between Two Bodies: Pouring Cold Water in a Hot Pan](#) if 0.260 kg of water was placed in the pan and 0.0100 kg of the water evaporated immediately, leaving the remainder to come to a common temperature with the pan?

Exercise:**Problem:**

In some countries, liquid nitrogen is used on dairy trucks instead of mechanical refrigerators. A 3.00-hour delivery trip requires 200 L of liquid nitrogen, which has a density of 808 kg/m^3 .

(a) Calculate the heat transfer necessary to evaporate this amount of liquid nitrogen and raise its temperature to 3.00°C. (Use c_p and assume it is constant over the temperature range.) This value is the amount of cooling the liquid nitrogen supplies.

(b) What is this heat transfer rate in kilowatt-hours?

(c) Compare the amount of cooling obtained from melting an identical mass of 0°C ice with that from evaporating the liquid nitrogen.

Solution:

(a) $1.57 \times 10^4\text{ kcal}$

(b) 18.3 kW · h

(c) $1.29 \times 10^4\text{ kcal}$

Exercise:**Problem:**

Some gun fanciers make their own bullets, which involves melting and casting the lead slugs. How much heat transfer is needed to raise the temperature and melt 0.500 kg of lead, starting from 25.0°C ?

Glossary

heat of sublimation

the energy required to change a substance from the solid phase to the vapor phase

latent heat coefficient

a physical constant equal to the amount of heat transferred for every 1 kg of a substance during the change in phase of the substance

sublimation

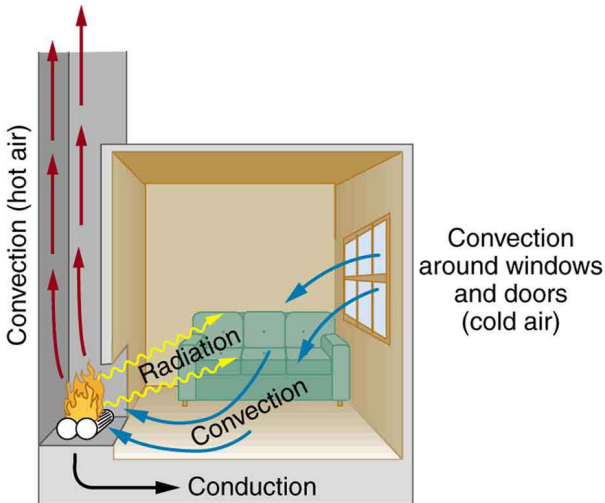
the transition from the solid phase to the vapor phase

Heat Transfer Methods

- Discuss the different methods of heat transfer.

Equally as interesting as the effects of heat transfer on a system are the methods by which this occurs. Whenever there is a temperature difference, heat transfer occurs. Heat transfer may occur rapidly, such as through a cooking pan, or slowly, such as through the walls of a picnic ice chest. We can control rates of heat transfer by choosing materials (such as thick wool clothing for the winter), controlling air movement (such as the use of weather stripping around doors), or by choice of color (such as a white roof to reflect summer sunlight). So many processes involve heat transfer, so that it is hard to imagine a situation where no heat transfer occurs. Yet every process involving heat transfer takes place by only three methods:

1. **Conduction** is heat transfer through stationary matter by physical contact. (The matter is stationary on a macroscopic scale—we know there is thermal motion of the atoms and molecules at any temperature above absolute zero.) Heat transferred between the electric burner of a stove and the bottom of a pan is transferred by conduction.
2. **Convection** is the heat transfer by the macroscopic movement of a fluid. This type of transfer takes place in a forced-air furnace and in weather systems, for example.
3. Heat transfer by **radiation** occurs when microwaves, infrared radiation, visible light, or another form of electromagnetic radiation is emitted or absorbed. An obvious example is the warming of the Earth by the Sun. A less obvious example is thermal radiation from the human body.



In a fireplace, heat transfer occurs by all three methods: conduction, convection, and radiation. Radiation is responsible for most of the heat transferred into the room. Heat transfer also occurs through conduction into the room, but at a much slower rate. Heat transfer by convection also occurs through cold air entering the room around windows and hot air leaving the room by rising up the chimney.

We examine these methods in some detail in the three following modules. Each method has unique and interesting characteristics, but all three do have one thing in common: they transfer heat solely because of a temperature difference [\[link\]](#).

Exercise:

Check Your Understanding

Problem:

Name an example from daily life (different from the text) for each mechanism of heat transfer.

Solution:

Conduction: Heat transfers into your hands as you hold a hot cup of coffee.

Convection: Heat transfers as the barista “steams” cold milk to make hot *cocoa*.

Radiation: Reheating a cold cup of coffee in a microwave oven.

Summary

- Heat is transferred by three different methods: conduction, convection, and radiation.

Conceptual Questions**Exercise:****Problem:**

What are the main methods of heat transfer from the hot core of Earth to its surface? From Earth’s surface to outer space?

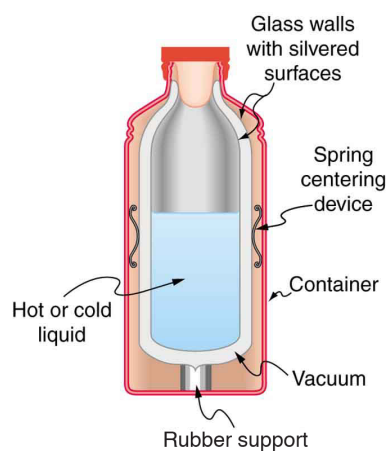
Exercise:**Problem:**

When our bodies get too warm, they respond by sweating and increasing blood circulation to the surface to transfer thermal energy away from the core. What effect will this have on a person in a 40.0°C hot tub?

Exercise:

Problem:

[\[link\]](#) shows a cut-away drawing of a thermos bottle (also known as a Dewar flask), which is a device designed specifically to slow down all forms of heat transfer. Explain the functions of the various parts, such as the vacuum, the silvering of the walls, the thin-walled long glass neck, the rubber support, the air layer, and the stopper.



The construction of a thermos bottle is designed to inhibit all methods of heat transfer.

Glossary

conduction

heat transfer through stationary matter by physical contact

convection

heat transfer by the macroscopic movement of fluid

radiation

heat transfer which occurs when microwaves, infrared radiation, visible light, or other electromagnetic radiation is emitted or absorbed

Conduction

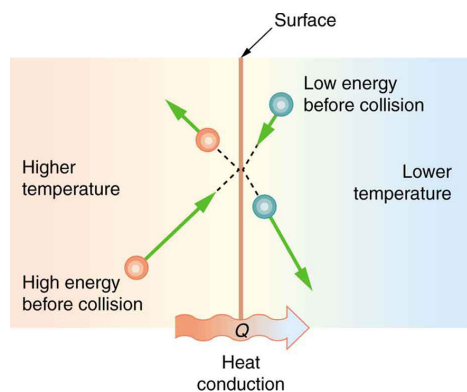
- Calculate thermal conductivity.
- Observe conduction of heat in collisions.
- Study thermal conductivities of common substances.



Insulation is used to limit the conduction of heat from the inside to the outside (in winters) and from the outside to the inside (in summers).
(credit: Giles Douglas)

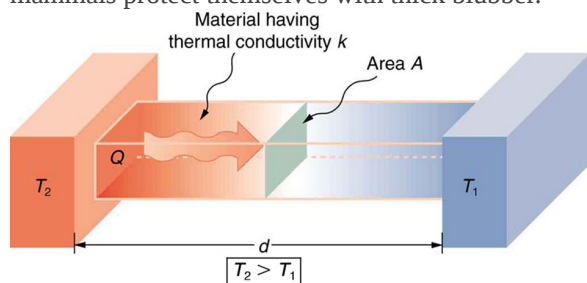
Your feet feel cold as you walk barefoot across the living room carpet in your cold house and then step onto the kitchen tile floor. This result is intriguing, since the carpet and tile floor are both at the same temperature. The different sensation you feel is explained by the different rates of heat transfer: the heat loss during the same time interval is greater for skin in contact with the tiles than with the carpet, so the temperature drop is greater on the tiles.

Some materials conduct thermal energy faster than others. In general, good conductors of electricity (metals like copper, aluminum, gold, and silver) are also good heat conductors, whereas insulators of electricity (wood, plastic, and rubber) are poor heat conductors. [\[link\]](#) shows molecules in two bodies at different temperatures. The (average) kinetic energy of a molecule in the hot body is higher than in the colder body. If two molecules collide, an energy transfer from the molecule with greater kinetic energy to the molecule with less kinetic energy occurs. The cumulative effect from all collisions results in a net flux of heat from the hot body to the colder body. The heat flux thus depends on the temperature difference $\Delta T = T_{\text{hot}} - T_{\text{cold}}$. Therefore, you will get a more severe burn from boiling water than from hot tap water. Conversely, if the temperatures are the same, the net heat transfer rate falls to zero, and equilibrium is achieved. Owing to the fact that the number of collisions increases with increasing area, heat conduction depends on the cross-sectional area. If you touch a cold wall with your palm, your hand cools faster than if you just touch it with your fingertip.



The molecules in two bodies at different temperatures have different average kinetic energies. Collisions occurring at the contact surface tend to transfer energy from high-temperature regions to low-temperature regions. In this illustration, a molecule in the lower temperature region (right side) has low energy before collision, but its energy increases after colliding with the contact surface. In contrast, a molecule in the higher temperature region (left side) has high energy before collision, but its energy decreases after colliding with the contact surface.

A third factor in the mechanism of conduction is the thickness of the material through which heat transfers. The figure below shows a slab of material with different temperatures on either side. Suppose that T_2 is greater than T_1 , so that heat is transferred from left to right. Heat transfer from the left side to the right side is accomplished by a series of molecular collisions. The thicker the material, the more time it takes to transfer the same amount of heat. This model explains why thick clothing is warmer than thin clothing in winters, and why Arctic mammals protect themselves with thick blubber.



Heat conduction occurs through any material, represented here by a rectangular bar, whether window glass or walrus blubber. The temperature of the material is T_2 on the left and T_1 on the right, where T_2 is greater than T_1 .

The rate of heat transfer by conduction is directly proportional to the surface area A , the temperature difference $T_2 - T_1$, and the substance's conductivity k . The rate of heat transfer is inversely proportional to the thickness d .

Lastly, the heat transfer rate depends on the material properties described by the coefficient of thermal conductivity. All four factors are included in a simple equation that was deduced from and is confirmed by experiments. The **rate of conductive heat transfer** through a slab of material, such as the one in [\[link\]](#), is given by

Equation:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d},$$

where Q/t is the rate of heat transfer in watts or kilocalories per second, k is the **thermal conductivity** of the material, A and d are its surface area and thickness, as shown in [\[link\]](#), and $(T_2 - T_1)$ is the temperature difference across the slab. [\[link\]](#) gives representative values of thermal conductivity.

Example:

Calculating Heat Transfer Through Conduction: Conduction Rate Through an Ice Box

A Styrofoam ice box has a total area of 0.950 m^2 and walls with an average thickness of 2.50 cm . The box contains ice, water, and canned beverages at 0°C . The inside of the box is kept cold by melting ice. How much ice melts in one day if the ice box is kept in the trunk of a car at 35.0°C ?

Strategy

This question involves both heat for a phase change (melting of ice) and the transfer of heat by conduction. To find the amount of ice melted, we must find the net heat transferred. This value can be obtained by calculating the rate of heat transfer by conduction and multiplying by time.

Solution

1. Identify the knowns.

Equation:

$$A = 0.950 \text{ m}^2; d = 2.50 \text{ cm} = 0.0250 \text{ m}; T_1 = 0^\circ\text{C}; T_2 = 35.0^\circ\text{C}; t = 1 \text{ day} = 24 \text{ hours} = 86,400 \text{ s}.$$

2. Identify the unknowns. We need to solve for the mass of the ice, m . We will also need to solve for the net heat transferred to melt the ice, Q .
3. Determine which equations to use. The rate of heat transfer by conduction is given by

Equation:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d}.$$

4. The heat is used to melt the ice: $Q = mL_f$.
5. Insert the known values:

Equation:

$$\frac{Q}{t} = \frac{(0.010 \text{ J/s} \cdot \text{m} \cdot ^\circ\text{C})(0.950 \text{ m}^2)(35.0^\circ\text{C} - 0^\circ\text{C})}{0.0250 \text{ m}} = 13.3 \text{ J/s}.$$

6. Multiply the rate of heat transfer by the time (1 day = 86,400 s):

Equation:

$$Q = (Q/t)t = (13.3 \text{ J/s})(86,400 \text{ s}) = 1.15 \times 10^6 \text{ J}.$$

7. Set this equal to the heat transferred to melt the ice: $Q = mL_f$. Solve for the mass m :

Equation:

$$m = \frac{Q}{L_f} = \frac{1.15 \times 10^6 \text{ J}}{334 \times 10^3 \text{ J/kg}} = 3.44 \text{ kg}.$$

Discussion

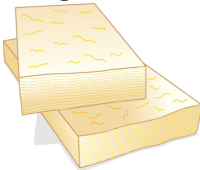
The result of 3.44 kg, or about 7.6 lbs, seems about right, based on experience. You might expect to use about a 4 kg (7–10 lb) bag of ice per day. A little extra ice is required if you add any warm food or beverages. Inspecting the conductivities in [\[link\]](#) shows that Styrofoam is a very poor conductor and thus a good insulator. Other good insulators include fiberglass, wool, and goose-down feathers. Like Styrofoam, these all incorporate many small pockets of air, taking advantage of air's poor thermal conductivity.

Substance	Thermal conductivity k (J/s·m·°C)
Silver	420
Copper	390
Gold	318
Aluminum	220
Steel iron	80
Steel (stainless)	14
Ice	2.2
Glass (average)	0.84
Concrete brick	0.84
Water	0.6
Fatty tissue (without blood)	0.2
Asbestos	0.16
Plasterboard	0.16
Wood	0.08–0.16

Substance	Thermal conductivity k (J/s·m·°C)
Snow (dry)	0.10
Cork	0.042
Glass wool	0.042
Wool	0.04
Down feathers	0.025
Air	0.023
Styrofoam	0.010

Thermal Conductivities of Common Substances[\[footnote\]](#)
At temperatures near 0°C.

A combination of material and thickness is often manipulated to develop good insulators—the smaller the conductivity k and the larger the thickness d , the better. The ratio of d/k will thus be large for a good insulator. The ratio d/k is called the **R factor**. The rate of conductive heat transfer is inversely proportional to R . The larger the value of R , the better the insulation. R factors are most commonly quoted for household insulation, refrigerators, and the like—unfortunately, it is still in non-metric units of $\text{ft}^2 \cdot ^\circ\text{F} \cdot \text{h} / \text{Btu}$, although the unit usually goes unstated (1 British thermal unit [Btu] is the amount of energy needed to change the temperature of 1.0 lb of water by 1.0 °F). A couple of representative values are an R factor of 11 for 3.5-in-thick fiberglass batts (pieces) of insulation and an R factor of 19 for 6.5-in-thick fiberglass batts. Walls are usually insulated with 3.5-in batts, while ceilings are usually insulated with 6.5-in batts. In cold climates, thicker batts may be used in ceilings and walls.



The fiberglass batt is used for insulation of walls and ceilings to prevent heat transfer between the inside of the building and the outside environment.

Note that in [\[link\]](#), the best thermal conductors—silver, copper, gold, and aluminum—are also the best electrical conductors, again related to the density of free electrons in them. Cooking utensils are typically made

from good conductors.

Example:

Calculating the Temperature Difference Maintained by a Heat Transfer: Conduction Through an Aluminum Pan

Water is boiling in an aluminum pan placed on an electrical element on a stovetop. The sauce pan has a bottom that is 0.800 cm thick and 14.0 cm in diameter. The boiling water is evaporating at the rate of 1.00 g/s. What is the temperature difference across (through) the bottom of the pan?

Strategy

Conduction through the aluminum is the primary method of heat transfer here, and so we use the equation for the rate of heat transfer and solve for the temperature difference.

Equation:

$$T_2 - T_1 = \frac{Q}{t} \left(\frac{d}{kA} \right).$$

Solution

1. Identify the knowns and convert them to the SI units.

The thickness of the pan, $d = 0.800 \text{ cm} = 8.0 \times 10^{-3} \text{ m}$, the area of the pan, $A = \pi(0.14/2)^2 \text{ m}^2 = 1.54 \times 10^{-2} \text{ m}^2$, and the thermal conductivity, $k = 220 \text{ J/s} \cdot \text{m} \cdot ^\circ\text{C}$.

2. Calculate the necessary heat of vaporization of 1 g of water:

Equation:

$$Q = mL_v = (1.00 \times 10^{-3} \text{ kg})(2256 \times 10^3 \text{ J/kg}) = 2256 \text{ J}.$$

3. Calculate the rate of heat transfer given that 1 g of water melts in one second:

Equation:

$$Q/t = 2256 \text{ J/s or } 2.26 \text{ kW}.$$

4. Insert the knowns into the equation and solve for the temperature difference:

Equation:

$$T_2 - T_1 = \frac{Q}{t} \left(\frac{d}{kA} \right) = (2256 \text{ J/s}) \frac{8.00 \times 10^{-3} \text{ m}}{(220 \text{ J/s} \cdot \text{m} \cdot ^\circ\text{C})(1.54 \times 10^{-2} \text{ m}^2)} = 5.33^\circ\text{C}.$$

Discussion

The value for the heat transfer $Q/t = 2.26 \text{ kW}$ or 2256 J/s is typical for an electric stove. This value gives a remarkably small temperature difference between the stove and the pan. Consider that the stove burner is red hot while the inside of the pan is nearly 100°C because of its contact with boiling water. This contact effectively cools the bottom of the pan in spite of its proximity to the very hot stove burner. Aluminum is such a good conductor that it only takes this small temperature difference to produce a heat transfer of 2.26 kW into the pan.

Conduction is caused by the random motion of atoms and molecules. As such, it is an ineffective mechanism for heat transport over macroscopic distances and short time distances. Take, for example, the temperature on the Earth, which would be unbearably cold during the night and extremely hot during the day if heat transport in the atmosphere was to be only through conduction. In another example, car engines would overheat unless there was a more efficient way to remove excess heat from the pistons.

Exercise:
Check Your Understanding

Problem:

How does the rate of heat transfer by conduction change when all spatial dimensions are doubled?

Solution:

Because area is the product of two spatial dimensions, it increases by a factor of four when each dimension is doubled ($A_{\text{final}} = (2d)^2 = 4d^2 = 4A_{\text{initial}}$). The distance, however, simply doubles. Because the temperature difference and the coefficient of thermal conductivity are independent of the spatial dimensions, the rate of heat transfer by conduction increases by a factor of four divided by two, or two:

Equation:

$$\left(\frac{Q}{t}\right)_{\text{final}} = \frac{kA_{\text{final}}(T_2 - T_1)}{d_{\text{final}}} = \frac{k(4A_{\text{initial}})(T_2 - T_1)}{2d_{\text{initial}}} = 2 \frac{kA_{\text{initial}}(T_2 - T_1)}{d_{\text{initial}}} = 2 \left(\frac{Q}{t}\right)_{\text{initial}}.$$

Summary

- Heat conduction is the transfer of heat between two objects in direct contact with each other.
- The rate of heat transfer Q/t (energy per unit time) is proportional to the temperature difference $T_2 - T_1$ and the contact area A and inversely proportional to the distance d between the objects:

Equation:

$$\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d}.$$

Conceptual Questions

Exercise:

Problem:

Some electric stoves have a flat ceramic surface with heating elements hidden beneath. A pot placed over a heating element will be heated, while it is safe to touch the surface only a few centimeters away. Why is ceramic, with a conductivity less than that of a metal but greater than that of a good insulator, an ideal choice for the stove top?

Exercise:

Problem:

Loose-fitting white clothing covering most of the body is ideal for desert dwellers, both in the hot Sun and during cold evenings. Explain how such clothing is advantageous during both day and night.



A jellabiya is worn by many men in Egypt. (credit: Zerida)

Problems & Exercises

Exercise:

Problem:

- (a) Calculate the rate of heat conduction through house walls that are 13.0 cm thick and that have an average thermal conductivity twice that of glass wool. Assume there are no windows or doors. The surface area of the walls is 120 m^2 and their inside surface is at 18.0°C , while their outside surface is at 5.00°C .
- (b) How many 1-kW room heaters would be needed to balance the heat transfer due to conduction?

Solution:

- (a) $1.01 \times 10^3 \text{ W}$
- (b) One

Exercise:

Problem:

The rate of heat conduction out of a window on a winter day is rapid enough to chill the air next to it. To see just how rapidly the windows transfer heat by conduction, calculate the rate of conduction in watts through a 3.00-m^2 window that is 0.635 cm thick (1/4 in) if the temperatures of the inner and outer surfaces are 5.00°C and -10.0°C , respectively. This rapid rate will not be maintained—the inner surface will cool, and even result in frost formation.

Exercise:

Problem:

Calculate the rate of heat conduction out of the human body, assuming that the core internal temperature is 37.0°C , the skin temperature is 34.0°C , the thickness of the tissues between averages 1.00 cm, and the surface area is 1.40 m^2 .

Solution:

84.0 W

Exercise:**Problem:**

Suppose you stand with one foot on ceramic flooring and one foot on a wool carpet, making contact over an area of 80.0 cm^2 with each foot. Both the ceramic and the carpet are 2.00 cm thick and are 10.0°C on their bottom sides. At what rate must heat transfer occur from each foot to keep the top of the ceramic and carpet at 33.0°C ?

Exercise:**Problem:**

A man consumes 3000 kcal of food in one day, converting most of it to maintain body temperature. If he loses half this energy by evaporating water (through breathing and sweating), how many kilograms of water evaporate?

Solution:

2.59 kg

Exercise:**Problem:**

(a) A firewalker runs across a bed of hot coals without sustaining burns. Calculate the heat transferred by conduction into the sole of one foot of a firewalker given that the bottom of the foot is a 3.00-mm-thick callus with a conductivity at the low end of the range for wood and its density is 300 kg/m^3 . The area of contact is 25.0 cm^2 , the temperature of the coals is 700°C , and the time in contact is 1.00 s.

(b) What temperature increase is produced in the 25.0 cm^3 of tissue affected?

(c) What effect do you think this will have on the tissue, keeping in mind that a callus is made of dead cells?

Exercise:**Problem:**

(a) What is the rate of heat conduction through the 3.00-cm-thick fur of a large animal having a 1.40-m^2 surface area? Assume that the animal's skin temperature is 32.0°C , that the air temperature is -5.00°C , and that fur has the same thermal conductivity as air. (b) What food intake will the animal need in one day to replace this heat transfer?

Solution:

(a) 39.7 W

(b) 820 kcal

Exercise:

Problem:

A walrus transfers energy by conduction through its blubber at the rate of 150 W when immersed in -1.00°C water. The walrus's internal core temperature is 37.0°C , and it has a surface area of 2.00 m^2 . What is the average thickness of its blubber, which has the conductivity of fatty tissues without blood?



Walrus on ice. (credit: Captain Budd Christman, NOAA Corps)

Exercise:**Problem:**

Compare the rate of heat conduction through a 13.0-cm-thick wall that has an area of 10.0 m^2 and a thermal conductivity twice that of glass wool with the rate of heat conduction through a window that is 0.750 cm thick and that has an area of 2.00 m^2 , assuming the same temperature difference across each.

Solution:

35 to 1, window to wall

Exercise:**Problem:**

Suppose a person is covered head to foot by wool clothing with average thickness of 2.00 cm and is transferring energy by conduction through the clothing at the rate of 50.0 W. What is the temperature difference across the clothing, given the surface area is 1.40 m^2 ?

Exercise:**Problem:**

Some stove tops are smooth ceramic for easy cleaning. If the ceramic is 0.600 cm thick and heat conduction occurs through the same area and at the same rate as computed in [\[link\]](#), what is the temperature difference across it? Ceramic has the same thermal conductivity as glass and brick.

Solution:

$1.05 \times 10^3\text{ K}$

Exercise:

Problem:

One easy way to reduce heating (and cooling) costs is to add extra insulation in the attic of a house. Suppose the house already had 15 cm of fiberglass insulation in the attic and in all the exterior surfaces. If you added an extra 8.0 cm of fiberglass to the attic, then by what percentage would the heating cost of the house drop? Take the single story house to be of dimensions 10 m by 15 m by 3.0 m. Ignore air infiltration and heat loss through windows and doors.

Exercise:**Problem:**

(a) Calculate the rate of heat conduction through a double-paned window that has a 1.50-m^2 area and is made of two panes of 0.800-cm-thick glass separated by a 1.00-cm air gap. The inside surface temperature is 15.0°C , while that on the outside is -10.0°C . (Hint: There are identical temperature drops across the two glass panes. First find these and then the temperature drop across the air gap. This problem ignores the increased heat transfer in the air gap due to convection.)

(b) Calculate the rate of heat conduction through a 1.60-cm-thick window of the same area and with the same temperatures. Compare your answer with that for part (a).

Solution:

(a) 83 W

(b) 24 times that of a double pane window.

Exercise:**Problem:**

Many decisions are made on the basis of the payback period: the time it will take through savings to equal the capital cost of an investment. Acceptable payback times depend upon the business or philosophy one has. (For some industries, a payback period is as small as two years.) Suppose you wish to install the extra insulation in [\[link\]](#). If energy cost \$1.00 per million joules and the insulation was \$4.00 per square meter, then calculate the simple payback time. Take the average ΔT for the 120 day heating season to be 15.0°C .

Exercise:**Problem:**

For the human body, what is the rate of heat transfer by conduction through the body's tissue with the following conditions: the tissue thickness is 3.00 cm, the change in temperature is 2.00°C , and the skin area is 1.50 m^2 . How does this compare with the average heat transfer rate to the body resulting from an energy intake of about 2400 kcal per day? (No exercise is included.)

Solution:

20.0 W, 17.2% of 2400 kcal per day

Glossary

R factor

the ratio of thickness to the conductivity of a material

rate of conductive heat transfer

rate of heat transfer from one material to another

thermal conductivity

the property of a material's ability to conduct heat

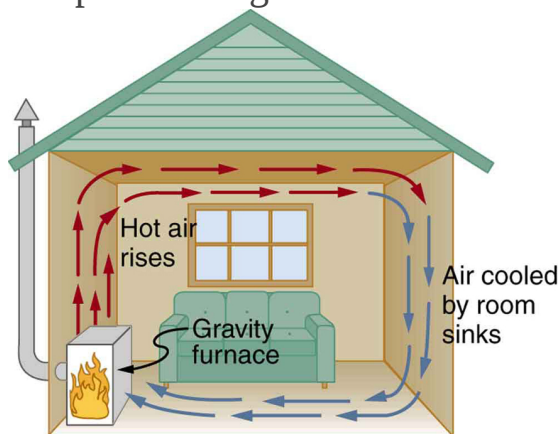
Convection

- Discuss the method of heat transfer by convection.

Convection is driven by large-scale flow of matter. In the case of Earth, the atmospheric circulation is caused by the flow of hot air from the tropics to the poles, and the flow of cold air from the poles toward the tropics. (Note that Earth's rotation causes the observed easterly flow of air in the northern hemisphere). Car engines are kept cool by the flow of water in the cooling system, with the water pump maintaining a flow of cool water to the pistons. The circulatory system is used the body: when the body overheats, the blood vessels in the skin expand (dilate), which increases the blood flow to the skin where it can be cooled by sweating. These vessels become smaller when it is cold outside and larger when it is hot (so more fluid flows, and more energy is transferred).

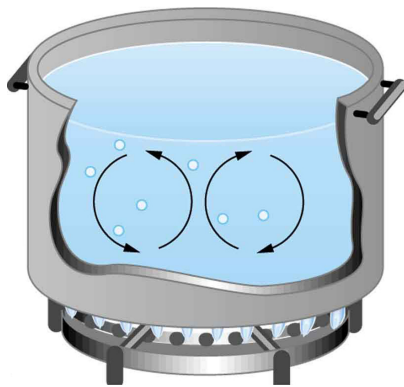
The body also loses a significant fraction of its heat through the breathing process.

While convection is usually more complicated than conduction, we can describe convection and do some straightforward, realistic calculations of its effects. Natural convection is driven by buoyant forces: hot air rises because density decreases as temperature increases. The house in [\[link\]](#) is kept warm in this manner, as is the pot of water on the stove in [\[link\]](#). Ocean currents and large-scale atmospheric circulation transfer energy from one part of the globe to another. Both are examples of natural convection.



Air heated by the so-called

gravity furnace expands and rises, forming a convective loop that transfers energy to other parts of the room. As the air is cooled at the ceiling and outside walls, it contracts, eventually becoming denser than room air and sinking to the floor. A properly designed heating system using natural convection, like this one, can be quite efficient in uniformly heating a home.



Convection plays an important role in heat transfer inside this pot of water. Once conducted to the inside, heat transfer to other parts of the pot is mostly by convection. The hotter water expands, decreases

in density, and rises to transfer heat to other regions of the water, while colder water sinks to the bottom. This process keeps repeating.

Note:

Take-Home Experiment: Convection Rolls in a Heated Pan

Take two small pots of water and use an eye dropper to place a drop of food coloring near the bottom of each. Leave one on a bench top and heat the other over a stovetop. Watch how the color spreads and how long it takes the color to reach the top. Watch how convective loops form.

Example:

Calculating Heat Transfer by Convection: Convection of Air Through the Walls of a House

Most houses are not airtight: air goes in and out around doors and windows, through cracks and crevices, following wiring to switches and outlets, and so on. The air in a typical house is completely replaced in less than an hour. Suppose that a moderately-sized house has inside dimensions $12.0\text{m} \times 18.0\text{m} \times 3.00\text{m}$ high, and that all air is replaced in 30.0 min. Calculate the heat transfer per unit time in watts needed to warm the incoming cold air by 10.0°C , thus replacing the heat transferred by convection alone.

Strategy

Heat is used to raise the temperature of air so that $Q = mc\Delta T$. The rate of heat transfer is then Q/t , where t is the time for air turnover. We are given that ΔT is 10.0°C , but we must still find values for the mass of air and its

specific heat before we can calculate Q . The specific heat of air is a weighted average of the specific heats of nitrogen and oxygen, which gives $c = c_p \cong 1000 \text{ J/kg} \cdot ^\circ \text{C}$ from [\[link\]](#) (note that the specific heat at constant pressure must be used for this process).

Solution

1. Determine the mass of air from its density and the given volume of the house. The density is given from the density ρ and the volume

Equation:

$$m = \rho V = (1.29 \text{ kg/m}^3)(12.0 \text{ m} \times 18.0 \text{ m} \times 3.00 \text{ m}) = 836 \text{ kg}.$$

2. Calculate the heat transferred from the change in air temperature:

$$Q = mc\Delta T \text{ so that}$$

Equation:

$$Q = (836 \text{ kg})(1000 \text{ J/kg} \cdot ^\circ \text{C})(10.0^\circ \text{C}) = 8.36 \times 10^6 \text{ J}.$$

3. Calculate the heat transfer from the heat Q and the turnover time t . Since air is turned over in $t = 0.500 \text{ h} = 1800 \text{ s}$, the heat transferred per unit time is

Equation:

$$\frac{Q}{t} = \frac{8.36 \times 10^6 \text{ J}}{1800 \text{ s}} = 4.64 \text{ kW}.$$

Discussion

This rate of heat transfer is equal to the power consumed by about forty-six 100-W light bulbs. Newly constructed homes are designed for a turnover time of 2 hours or more, rather than 30 minutes for the house of this example. Weather stripping, caulking, and improved window seals are commonly employed. More extreme measures are sometimes taken in very cold (or hot) climates to achieve a tight standard of more than 6 hours for one air turnover. Still longer turnover times are unhealthy, because a minimum amount of fresh air is necessary to supply oxygen for breathing and to dilute household pollutants. The term used for the process by which

outside air leaks into the house from cracks around windows, doors, and the foundation is called “air infiltration.”

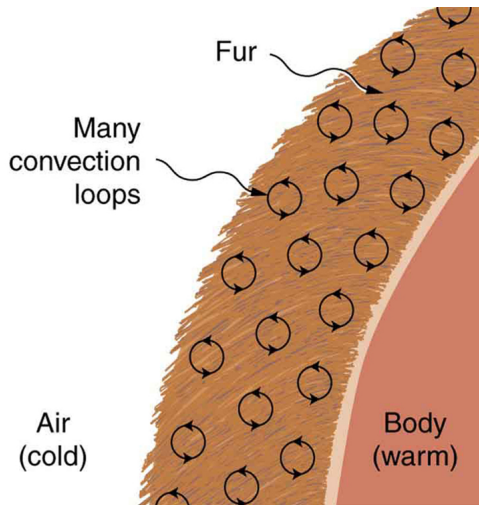
A cold wind is much more chilling than still cold air, because convection combines with conduction in the body to increase the rate at which energy is transferred away from the body. The table below gives approximate wind-chill factors, which are the temperatures of still air that produce the same rate of cooling as air of a given temperature and speed. Wind-chill factors are a dramatic reminder of convection’s ability to transfer heat faster than conduction. For example, a 15.0 m/s wind at 0°C has the chilling equivalent of still air at about -18°C .

Moving air temperature	Wind speed (m/s)				
(°C)	2	5	10	15	20
5	3	−1	−8	−10	−12
2	0	−7	−12	−16	−18
0	−2	−9	−15	−18	−20

Moving air temperature	Wind speed (m/s)				
−5	−7	−15	−22	−26	−29
−10	−12	−21	−29	−34	−36
−20	−23	−34	−44	−50	−52
−40	−44	−59	−73	−82	−84

Wind-Chill Factors

Although air can transfer heat rapidly by convection, it is a poor conductor and thus a good insulator. The amount of available space for airflow determines whether air acts as an insulator or conductor. The space between the inside and outside walls of a house, for example, is about 9 cm (3.5 in)—large enough for convection to work effectively. The addition of wall insulation prevents airflow, so heat loss (or gain) is decreased. Similarly, the gap between the two panes of a double-paned window is about 1 cm, which prevents convection and takes advantage of air's low conductivity to prevent greater loss. Fur, fiber, and fiberglass also take advantage of the low conductivity of air by trapping it in spaces too small to support convection, as shown in the figure. Fur and feathers are lightweight and thus ideal for the protection of animals.



Fur is filled with air, breaking it up into many small pockets. Convection is very slow here, because the loops are so small. The low conductivity of air makes fur a very good lightweight insulator.

Some interesting phenomena happen *when convection is accompanied by a phase change*. It allows us to cool off by sweating, even if the temperature of the surrounding air exceeds body temperature. Heat from the skin is required for sweat to evaporate from the skin, but without air flow, the air becomes saturated and evaporation stops. Air flow caused by convection replaces the saturated air by dry air and evaporation continues.

Example:

Calculate the Flow of Mass during Convection: Sweat-Heat Transfer away from the Body

The average person produces heat at the rate of about 120 W when at rest. At what rate must water evaporate from the body to get rid of all this energy? (This evaporation might occur when a person is sitting in the shade and surrounding temperatures are the same as skin temperature, eliminating heat transfer by other methods.)

Strategy

Energy is needed for a phase change ($Q = mL_v$). Thus, the energy loss per unit time is

Equation:

$$\frac{Q}{t} = \frac{mL_v}{t} = 120 \text{ W} = 120 \text{ J/s}.$$

We divide both sides of the equation by L_v to find that the mass evaporated per unit time is

Equation:

$$\frac{m}{t} = \frac{120 \text{ J/s}}{L_v}.$$

Solution

(1) Insert the value of the latent heat from [\[link\]](#), $L_v = 2430 \text{ kJ/kg} = 2430 \text{ J/g}$. This yields

Equation:

$$\frac{m}{t} = \frac{120 \text{ J/s}}{2430 \text{ J/g}} = 0.0494 \text{ g/s} = 2.96 \text{ g/min}.$$

Discussion

Evaporating about 3 g/min seems reasonable. This would be about 180 g (about 7 oz) per hour. If the air is very dry, the sweat may evaporate without even being noticed. A significant amount of evaporation also takes place in the lungs and breathing passages.

Another important example of the combination of phase change and convection occurs when water evaporates from the oceans. Heat is removed

from the ocean when water evaporates. If the water vapor condenses in liquid droplets as clouds form, heat is released in the atmosphere. Thus, there is an overall transfer of heat from the ocean to the atmosphere. This process is the driving power behind thunderheads, those great cumulus clouds that rise as much as 20.0 km into the stratosphere. Water vapor carried in by convection condenses, releasing tremendous amounts of energy. This energy causes the air to expand and rise, where it is colder. More condensation occurs in these colder regions, which in turn drives the cloud even higher. Such a mechanism is called positive feedback, since the process reinforces and accelerates itself. These systems sometimes produce violent storms, with lightning and hail, and constitute the mechanism driving hurricanes.



Cumulus clouds
are caused by
water vapor that
rises because of
convection. The
rise of clouds is
driven by a
positive
feedback
mechanism.
(credit: Mike
Love)



Convection
accompanied by a
phase change
releases the energy
needed to drive this
thunderhead into the
stratosphere. (credit:
Gerardo García
Moretti)



The phase change that
occurs when this
iceberg melts involves
tremendous heat

transfer. (credit:
Dominic Alves)

The movement of icebergs is another example of convection accompanied by a phase change. Suppose an iceberg drifts from Greenland into warmer Atlantic waters. Heat is removed from the warm ocean water when the ice melts and heat is released to the land mass when the iceberg forms on Greenland.

Exercise:

Check Your Understanding

Problem: Explain why using a fan in the summer feels refreshing!

Solution:

Using a fan increases the flow of air: warm air near your body is replaced by cooler air from elsewhere. Convection increases the rate of heat transfer so that moving air “feels” cooler than still air.

Summary

- Convection is heat transfer by the macroscopic movement of mass. Convection can be natural or forced and generally transfers thermal energy faster than conduction. [\[link\]](#) gives wind-chill factors, indicating that moving air has the same chilling effect of much colder stationary air. *Convection that occurs along with a phase change* can transfer energy from cold regions to warm ones.

Conceptual Questions

Exercise:

Problem:

One way to make a fireplace more energy efficient is to have an external air supply for the combustion of its fuel. Another is to have room air circulate around the outside of the fire box and back into the room. Detail the methods of heat transfer involved in each.

Exercise:**Problem:**

On cold, clear nights horses will sleep under the cover of large trees. How does this help them keep warm?

Problems & Exercises**Exercise:****Problem:**

At what wind speed does -10°C air cause the same chill factor as still air at -29°C ?

Solution:

10 m/s

Exercise:**Problem:**

At what temperature does still air cause the same chill factor as -5°C air moving at 15 m/s?

Exercise:

Problem:

The “steam” above a freshly made cup of instant coffee is really water vapor droplets condensing after evaporating from the hot coffee. What is the final temperature of 250 g of hot coffee initially at 90.0°C if 2.00 g evaporates from it? The coffee is in a Styrofoam cup, so other methods of heat transfer can be neglected.

Solution:

85.7°C

Exercise:**Problem:**

(a) How many kilograms of water must evaporate from a 60.0-kg woman to lower her body temperature by 0.750°C ?

(b) Is this a reasonable amount of water to evaporate in the form of perspiration, assuming the relative humidity of the surrounding air is low?

Exercise:**Problem:**

On a hot dry day, evaporation from a lake has just enough heat transfer to balance the $1.00\text{ kW}/\text{m}^2$ of incoming heat from the Sun. What mass of water evaporates in 1.00 h from each square meter? Explicitly show how you follow the steps in the [Problem-Solving Strategies for the Effects of Heat Transfer](#).

Solution:

1.48 kg

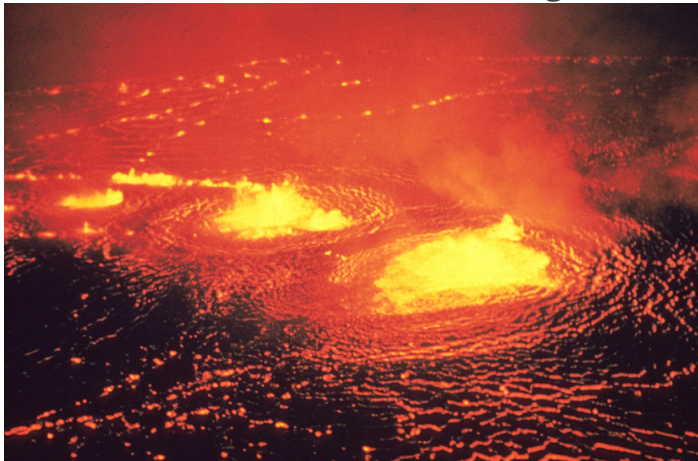
Exercise:

Problem:

One winter day, the climate control system of a large university classroom building malfunctions. As a result, 500 m^3 of excess cold air is brought in each minute. At what rate in kilowatts must heat transfer occur to warm this air by 10.0°C (that is, to bring the air to room temperature)?

Exercise:**Problem:**

The Kilauea volcano in Hawaii is the world's most active, disgorging about $5 \times 10^5 \text{ m}^3$ of 1200°C lava per day. What is the rate of heat transfer out of Earth by convection if this lava has a density of 2700 kg/m^3 and eventually cools to 30°C ? Assume that the specific heat of lava is the same as that of granite.



Lava flow on Kilauea volcano in Hawaii. (credit: J. P. Eaton, U.S. Geological Survey)

Solution:

$$2 \times 10^4 \text{ MW}$$

Exercise:**Problem:**

During heavy exercise, the body pumps 2.00 L of blood per minute to the surface, where it is cooled by 2.00°C . What is the rate of heat transfer from this forced convection alone, assuming blood has the same specific heat as water and its density is 1050 kg/m^3 ?

Exercise:**Problem:**

A person inhales and exhales 2.00 L of 37.0°C air, evaporating $4.00 \times 10^{-2}\text{ g}$ of water from the lungs and breathing passages with each breath.

- (a) How much heat transfer occurs due to evaporation in each breath?
- (b) What is the rate of heat transfer in watts if the person is breathing at a moderate rate of 18.0 breaths per minute?
- (c) If the inhaled air had a temperature of 20.0°C , what is the rate of heat transfer for warming the air?
- (d) Discuss the total rate of heat transfer as it relates to typical metabolic rates. Will this breathing be a major form of heat transfer for this person?

Solution:

- (a) 97.2 J
- (b) 29.2 W
- (c) 9.49 W
- (d) The total rate of heat loss would be $29.2\text{ W} + 9.49\text{ W} = 38.7\text{ W}$. While sleeping, our body consumes 83 W of power, while sitting it

consumes 120 to 210 W. Therefore, the total rate of heat loss from breathing will not be a major form of heat loss for this person.

Exercise:

Problem:

A glass coffee pot has a circular bottom with a 9.00-cm diameter in contact with a heating element that keeps the coffee warm with a continuous heat transfer rate of 50.0 W

(a) What is the temperature of the bottom of the pot, if it is 3.00 mm thick and the inside temperature is 60.0°C?

(b) If the temperature of the coffee remains constant and all of the heat transfer is removed by evaporation, how many grams per minute evaporate? Take the heat of vaporization to be 2340 kJ/kg.

Radiation

- Discuss heat transfer by radiation.
- Explain the power of different materials.

You can feel the heat transfer from a fire and from the Sun. Similarly, you can sometimes tell that the oven is hot without touching its door or looking inside—it may just warm you as you walk by. The space between the Earth and the Sun is largely empty, without any possibility of heat transfer by convection or conduction. In these examples, heat is transferred by radiation. That is, the hot body emits electromagnetic waves that are absorbed by our skin: no medium is required for electromagnetic waves to propagate. Different names are used for electromagnetic waves of different wavelengths: radio waves, microwaves, infrared **radiation**, visible light, ultraviolet radiation, X-rays, and gamma rays.

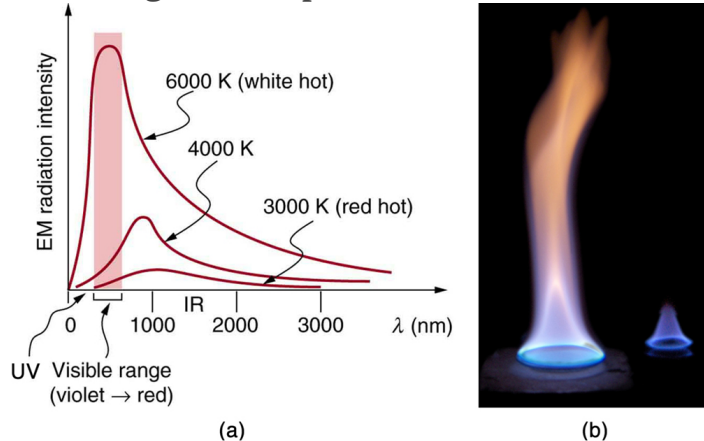


Most of the heat transfer from this fire to the observers is through infrared radiation. The visible light, although dramatic, transfers relatively little thermal energy. Convection transfers energy away from the observers as hot air rises, while conduction is negligibly slow here. Skin is very sensitive to infrared radiation, so that you can sense the presence of a fire

without looking at it directly. (credit:
Daniel X. O'Neil)

The energy of electromagnetic radiation depends on the wavelength (color) and varies over a wide range: a smaller wavelength (or higher frequency) corresponds to a higher energy. Because more heat is radiated at higher temperatures, a temperature change is accompanied by a color change. Take, for example, an electrical element on a stove, which glows from red to orange, while the higher-temperature steel in a blast furnace glows from yellow to white. The radiation you feel is mostly infrared, which corresponds to a lower temperature than that of the electrical element and the steel. The radiated energy depends on its intensity, which is represented in the figure below by the height of the distribution.

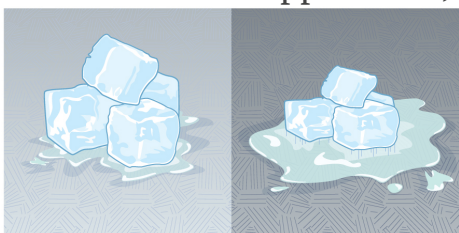
[Electromagnetic Waves](#) explains more about the electromagnetic spectrum and [Introduction to Quantum Physics](#) discusses how the decrease in wavelength corresponds to an increase in energy.



(a) A graph of the spectra of electromagnetic waves emitted from an ideal radiator at three different temperatures. The intensity or rate of radiation emission increases dramatically with temperature, and the spectrum shifts toward the visible and ultraviolet parts of the spectrum. The

shaded portion denotes the visible part of the spectrum. It is apparent that the shift toward the ultraviolet with temperature makes the visible appearance shift from red to white to blue as temperature increases. (b)
Note the variations in color corresponding to variations in flame temperature. (credit: Tuohirulla)

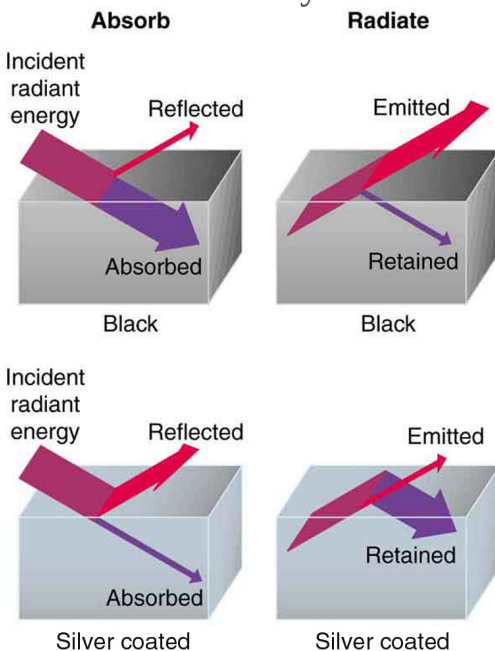
All objects absorb and emit electromagnetic radiation. The rate of heat transfer by radiation is largely determined by the color of the object. Black is the most effective, and white is the least effective. People living in hot climates generally avoid wearing black clothing, for instance (see [\[link\]](#)). Similarly, black asphalt in a parking lot will be hotter than adjacent gray sidewalk on a summer day, because black absorbs better than gray. The reverse is also true—black radiates better than gray. Thus, on a clear summer night, the asphalt will be colder than the gray sidewalk, because black radiates the energy more rapidly than gray. An *ideal radiator* is the same color as an *ideal absorber*, and captures all the radiation that falls on it. In contrast, white is a poor absorber and is also a poor radiator. A white object reflects all radiation, like a mirror. (A perfect, polished white surface is mirror-like in appearance, and a crushed mirror looks white.)



This illustration shows that the darker pavement is hotter than the lighter pavement (much more of the ice on the right has

melted), although both have been in the sunlight for the same time. The thermal conductivities of the pavements are the same.

Gray objects have a uniform ability to absorb all parts of the electromagnetic spectrum. Colored objects behave in similar but more complex ways, which gives them a particular color in the visible range and may make them special in other ranges of the nonvisible spectrum. Take, for example, the strong absorption of infrared radiation by the skin, which allows us to be very sensitive to it.



A black object is a good absorber and a good radiator, while a white (or silver) object is a poor absorber and a poor radiator. It is as if

radiation from the inside is reflected back into the silver object, whereas radiation from the inside of the black object is “absorbed” when it hits the surface and finds itself on the outside and is strongly emitted.

The rate of heat transfer by emitted radiation is determined by the **Stefan-Boltzmann law of radiation**:

Equation:

$$\frac{Q}{t} = \sigma e A T^4,$$

where $\sigma = 5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4$ is the Stefan-Boltzmann constant, A is the surface area of the object, and T is its absolute temperature in kelvin. The symbol e stands for the **emissivity** of the object, which is a measure of how well it radiates. An ideal jet-black (or black body) radiator has $e = 1$, whereas a perfect reflector has $e = 0$. Real objects fall between these two values. Take, for example, tungsten light bulb filaments which have an e of about 0.5, and carbon black (a material used in printer toner), which has the (greatest known) emissivity of about 0.99.

The radiation rate is directly proportional to the *fourth power* of the absolute temperature—a remarkably strong temperature dependence. Furthermore, the radiated heat is proportional to the surface area of the object. If you knock apart the coals of a fire, there is a noticeable increase in radiation due to an increase in radiating surface area.



A thermograph of part of a building shows temperature variations, indicating where heat transfer to the outside is most severe. Windows are a major region of heat transfer to the outside of homes. (credit: U.S. Army)

Skin is a remarkably good absorber and emitter of infrared radiation, having an emissivity of 0.97 in the infrared spectrum. Thus, we are all nearly (jet) black in the infrared, in spite of the obvious variations in skin color. This high infrared emissivity is why we can so easily feel radiation on our skin. It is also the basis for the use of night scopes used by law enforcement and the military to detect human beings. Even small temperature variations can be detected because of the T^4 dependence. Images, called *thermographs*, can be used medically to detect regions of abnormally high temperature in the body, perhaps indicative of disease. Similar techniques can be used to detect heat leaks in homes [\[link\]](#), optimize performance of blast furnaces, improve comfort levels in work environments, and even remotely map the Earth's temperature profile.

All objects emit and absorb radiation. The *net* rate of heat transfer by radiation (absorption minus emission) is related to both the temperature of the object and the temperature of its surroundings. Assuming that an object

with a temperature T_1 is surrounded by an environment with uniform temperature T_2 , the **net rate of heat transfer by radiation** is

Equation:

$$\frac{Q_{\text{net}}}{t} = \sigma e A (T_2^4 - T_1^4),$$

where e is the emissivity of the object alone. In other words, it does not matter whether the surroundings are white, gray, or black; the balance of radiation into and out of the object depends on how well it emits and absorbs radiation. When $T_2 > T_1$, the quantity Q_{net}/t is positive; that is, the net heat transfer is from hot to cold.

Note:

Take-Home Experiment: Temperature in the Sun

Place a thermometer out in the sunshine and shield it from direct sunlight using an aluminum foil. What is the reading? Now remove the shield, and note what the thermometer reads. Take a handkerchief soaked in nail polish remover, wrap it around the thermometer and place it in the sunshine. What does the thermometer read?

Example:

Calculate the Net Heat Transfer of a Person: Heat Transfer by Radiation

What is the rate of heat transfer by radiation, with an unclothed person standing in a dark room whose ambient temperature is 22.0°C . The person has a normal skin temperature of 33.0°C and a surface area of 1.50 m^2 . The emissivity of skin is 0.97 in the infrared, where the radiation takes place.

Strategy

We can solve this by using the equation for the rate of radiative heat transfer.

Solution

Insert the temperatures values $T_2 = 295 \text{ K}$ and $T_1 = 306 \text{ K}$, so that

Equation:

$$\frac{Q}{t} = \sigma e A (T_2^4 - T_1^4)$$

Equation:

$$= (5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4)(0.97)(1.50 \text{ m}^2) \left[(295 \text{ K})^4 - (306 \text{ K})^4 \right]$$

Equation:

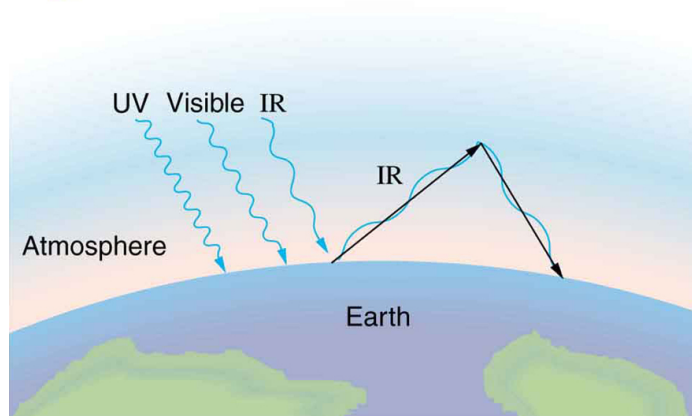
$$= -99 \text{ J/s} = -99 \text{ W}.$$

Discussion

This value is a significant rate of heat transfer to the environment (note the minus sign), considering that a person at rest may produce energy at the rate of 125 W and that conduction and convection will also be transferring energy to the environment. Indeed, we would probably expect this person to feel cold. Clothing significantly reduces heat transfer to the environment by many methods, because clothing slows down both conduction and convection, and has a lower emissivity (especially if it is white) than skin.

The Earth receives almost all its energy from radiation of the Sun and reflects some of it back into outer space. Because the Sun is hotter than the Earth, the net energy flux is from the Sun to the Earth. However, the rate of energy transfer is less than the equation for the radiative heat transfer would predict because the Sun does not fill the sky. The average emissivity (e) of the Earth is about 0.65, but the calculation of this value is complicated by the fact that the highly reflective cloud coverage varies greatly from day to day. There is a negative feedback (one in which a change produces an effect that opposes that change) between clouds and heat transfer; greater temperatures evaporate more water to form more clouds, which reflect more radiation back into space, reducing the temperature. The often mentioned **greenhouse effect** is directly related to the variation of the Earth's emissivity with radiation type (see the figure given below). The greenhouse

effect is a natural phenomenon responsible for providing temperatures suitable for life on Earth. The Earth's relatively constant temperature is a result of the energy balance between the incoming solar radiation and the energy radiated from the Earth. Most of the infrared radiation emitted from the Earth is absorbed by carbon dioxide (CO_2) and water (H_2O) in the atmosphere and then re-radiated back to the Earth or into outer space. Re-radiation back to the Earth maintains its surface temperature about 40°C higher than it would be if there was no atmosphere, similar to the way glass increases temperatures in a greenhouse.

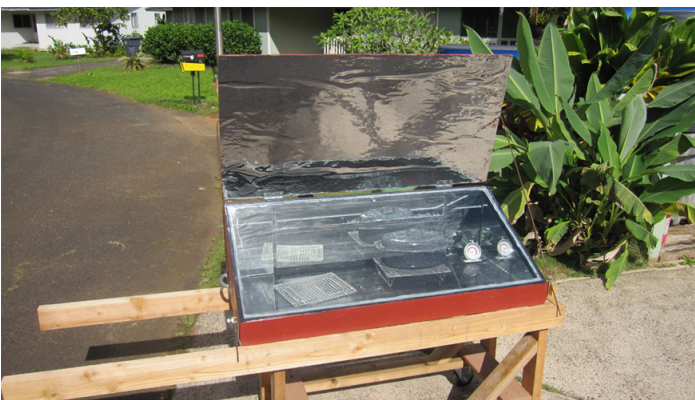


The greenhouse effect is a name given to the trapping of energy in the Earth's atmosphere by a process similar to that used in greenhouses. The atmosphere, like window glass, is transparent to incoming visible radiation and most of the Sun's infrared. These wavelengths are absorbed by the Earth and re-emitted as infrared. Since Earth's temperature is much lower than that of the Sun, the infrared radiated by the Earth has a much longer wavelength. The atmosphere, like glass, traps these longer infrared rays, keeping the Earth warmer than it would otherwise

be. The amount of trapping depends on concentrations of trace gases like carbon dioxide, and a change in the concentration of these gases is believed to affect the Earth's surface temperature.

The greenhouse effect is also central to the discussion of global warming due to emission of carbon dioxide and methane (and other so-called greenhouse gases) into the Earth's atmosphere from industrial production and farming. Changes in global climate could lead to more intense storms, precipitation changes (affecting agriculture), reduction in rain forest biodiversity, and rising sea levels.

Heating and cooling are often significant contributors to energy use in individual homes. Current research efforts into developing environmentally friendly homes quite often focus on reducing conventional heating and cooling through better building materials, strategically positioning windows to optimize radiation gain from the Sun, and opening spaces to allow convection. It is possible to build a zero-energy house that allows for comfortable living in most parts of the United States with hot and humid summers and cold winters.



This simple but effective solar cooker uses the greenhouse effect and reflective material to trap and retain solar energy. Made of inexpensive,

durable materials, it saves money and labor, and is of particular economic value in energy-poor developing countries. (credit: E.B. Kauai)

Conversely, dark space is very cold, about 3K (-454°F), so that the Earth radiates energy into the dark sky. Owing to the fact that clouds have lower emissivity than either oceans or land masses, they reflect some of the radiation back to the surface, greatly reducing heat transfer into dark space, just as they greatly reduce heat transfer into the atmosphere during the day. The rate of heat transfer from soil and grasses can be so rapid that frost may occur on clear summer evenings, even in warm latitudes.

Exercise:

Check Your Understanding

Problem:

What is the change in the rate of the radiated heat by a body at the temperature $T_1 = 20^{\circ}\text{C}$ compared to when the body is at the temperature $T_2 = 40^{\circ}\text{C}$?

Solution:

The radiated heat is proportional to the fourth power of the *absolute temperature*. Because $T_1 = 293\text{ K}$ and $T_2 = 313\text{ K}$, the rate of heat transfer increases by about 30 percent of the original rate.

Note:

Career Connection: Energy Conservation Consultation

The cost of energy is generally believed to remain very high for the foreseeable future. Thus, passive control of heat loss in both commercial and domestic housing will become increasingly important. Energy consultants measure and analyze the flow of energy into and out of houses

and ensure that a healthy exchange of air is maintained inside the house. The job prospects for an energy consultant are strong.

Note:

Problem-Solving Strategies for the Methods of Heat Transfer

1. *Examine the situation to determine what type of heat transfer is involved.*
2. *Identify the type(s) of heat transfer—conduction, convection, or radiation.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is very useful.*
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).*
6. For conduction, equation $\frac{Q}{t} = \frac{kA(T_2 - T_1)}{d}$ is appropriate. [\[link\]](#) lists thermal conductivities. For convection, determine the amount of matter moved and use equation $Q = mc\Delta T$, to calculate the heat transfer involved in the temperature change of the fluid. If a phase change accompanies convection, equation $Q = mL_f$ or $Q = mL_v$ is appropriate to find the heat transfer involved in the phase change. [\[link\]](#) lists information relevant to phase change. For radiation, equation $\frac{Q_{\text{net}}}{t} = \sigma eA(T_2^4 - T_1^4)$ gives the net heat transfer rate.
7. *Insert the knowns along with their units into the appropriate equation and obtain numerical solutions complete with units.*
8. *Check the answer to see if it is reasonable. Does it make sense?*

Summary

- Radiation is the rate of heat transfer through the emission or absorption of electromagnetic waves.

- The rate of heat transfer depends on the surface area and the fourth power of the absolute temperature:

Equation:

$$\frac{Q}{t} = \sigma e A T^4,$$

where $\sigma = 5.67 \times 10^{-8} \text{ J/s} \cdot \text{m}^2 \cdot \text{K}^4$ is the Stefan-Boltzmann constant and e is the emissivity of the body. For a black body, $e = 1$ whereas a shiny white or perfect reflector has $e = 0$, with real objects having values of e between 1 and 0. The net rate of heat transfer by radiation is

Equation:

$$\frac{Q_{\text{net}}}{t} = \sigma e A (T_2^4 - T_1^4)$$

where T_1 is the temperature of an object surrounded by an environment with uniform temperature T_2 and e is the emissivity of the *object*.

Conceptual Questions

Exercise:

Problem:

When watching a daytime circus in a large, dark-colored tent, you sense significant heat transfer from the tent. Explain why this occurs.

Exercise:

Problem:

Satellites designed to observe the radiation from cold (3 K) dark space have sensors that are shaded from the Sun, Earth, and Moon and that are cooled to very low temperatures. Why must the sensors be at low temperature?

Exercise:

Problem: Why are cloudy nights generally warmer than clear ones?

Exercise:**Problem:**

Why are thermometers that are used in weather stations shielded from the sunshine? What does a thermometer measure if it is shielded from the sunshine and also if it is not?

Exercise:**Problem:**

On average, would Earth be warmer or cooler without the atmosphere? Explain your answer.

Problems & Exercises**Exercise:****Problem:**

At what net rate does heat radiate from a 275-m^2 black roof on a night when the roof's temperature is 30.0°C and the surrounding temperature is 15.0°C ? The emissivity of the roof is 0.900.

Solution:

-21.7 kW

Note that the negative answer implies heat loss to the surroundings.

Exercise:

Problem:

(a) Cherry-red embers in a fireplace are at 850°C and have an exposed area of 0.200 m^2 and an emissivity of 0.980. The surrounding room has a temperature of 18.0°C . If 50% of the radiant energy enters the room, what is the net rate of radiant heat transfer in kilowatts? (b) Does your answer support the contention that most of the heat transfer into a room by a fireplace comes from infrared radiation?

Exercise:**Problem:**

Radiation makes it impossible to stand close to a hot lava flow. Calculate the rate of heat transfer by radiation from 1.00 m^2 of 1200°C fresh lava into 30.0°C surroundings, assuming lava's emissivity is 1.00.

Solution:

−266 kW

Exercise:**Problem:**

(a) Calculate the rate of heat transfer by radiation from a car radiator at 110°C into a 50.0°C environment, if the radiator has an emissivity of 0.750 and a 1.20-m^2 surface area. (b) Is this a significant fraction of the heat transfer by an automobile engine? To answer this, assume a horsepower of 200 hp (1.5 kW) and the efficiency of automobile engines as 25%.

Exercise:

Problem:

Find the net rate of heat transfer by radiation from a skier standing in the shade, given the following. She is completely clothed in white (head to foot, including a ski mask), the clothes have an emissivity of 0.200 and a surface temperature of 10.0°C , the surroundings are at -15.0°C , and her surface area is 1.60 m^2 .

Solution:

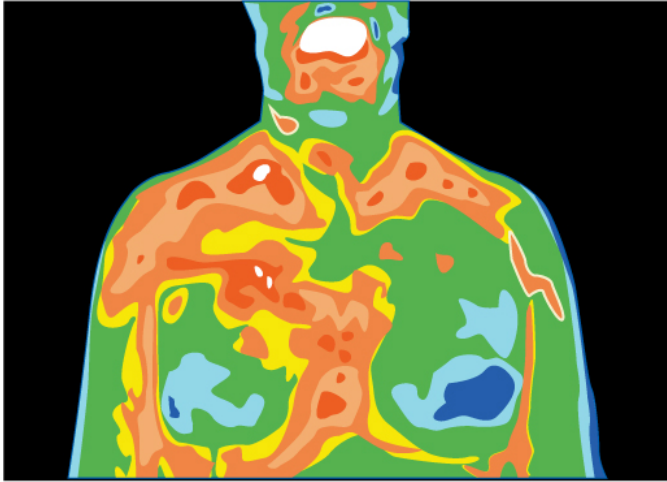
-36.0 W

Exercise:**Problem:**

Suppose you walk into a sauna that has an ambient temperature of 50.0°C . (a) Calculate the rate of heat transfer to you by radiation given your skin temperature is 37.0°C , the emissivity of skin is 0.98, and the surface area of your body is 1.50 m^2 . (b) If all other forms of heat transfer are balanced (the net heat transfer is zero), at what rate will your body temperature increase if your mass is 75.0 kg ?

Exercise:**Problem:**

Thermography is a technique for measuring radiant heat and detecting variations in surface temperatures that may be medically, environmentally, or militarily meaningful. (a) What is the percent increase in the rate of heat transfer by radiation from a given area at a temperature of 34.0°C compared with that at 33.0°C , such as on a person's skin? (b) What is the percent increase in the rate of heat transfer by radiation from a given area at a temperature of 34.0°C compared with that at 20.0°C , such as for warm and cool automobile hoods?



Artist's rendition of a thermograph of a patient's upper body, showing the distribution of heat represented by different colors.

Solution:

(a) 1.31%

(b) 20.5%

Exercise:

Problem:

The Sun radiates like a perfect black body with an emissivity of exactly 1. (a) Calculate the surface temperature of the Sun, given that it is a sphere with a 7.00×10^8 -m radius that radiates 3.80×10^{26} W into 3-K space. (b) How much power does the Sun radiate per square meter of its surface? (c) How much power in watts per square meter is that value at the distance of Earth, 1.50×10^{11} m away? (This number is called the solar constant.)

Exercise:

Problem:

A large body of lava from a volcano has stopped flowing and is slowly cooling. The interior of the lava is at 1200°C , its surface is at 450°C , and the surroundings are at 27.0°C . (a) Calculate the rate at which energy is transferred by radiation from 1.00 m^2 of surface lava into the surroundings, assuming the emissivity is 1.00. (b) Suppose heat conduction to the surface occurs at the same rate. What is the thickness of the lava between the 450°C surface and the 1200°C interior, assuming that the lava's conductivity is the same as that of brick?

Solution:

(a) -15.0 kW

(b) 4.2 cm

Exercise:**Problem:**

Calculate the temperature the entire sky would have to be in order to transfer energy by radiation at 1000 W/m^2 —about the rate at which the Sun radiates when it is directly overhead on a clear day. This value is the effective temperature of the sky, a kind of average that takes account of the fact that the Sun occupies only a small part of the sky but is much hotter than the rest. Assume that the body receiving the energy has a temperature of 27.0°C .

Exercise:

Problem:

(a) A shirtless rider under a circus tent feels the heat radiating from the sunlit portion of the tent. Calculate the temperature of the tent canvas based on the following information: The shirtless rider's skin temperature is 34.0°C and has an emissivity of 0.970. The exposed area of skin is 0.400 m^2 . He receives radiation at the rate of 20.0 W —half what you would calculate if the entire region behind him was hot. The rest of the surroundings are at 34.0°C . (b) Discuss how this situation would change if the sunlit side of the tent was nearly pure white and if the rider was covered by a white tunic.

Solution:

(a) 48.5°C

(b) A pure white object reflects more of the radiant energy that hits it, so a white tent would prevent more of the sunlight from heating up the inside of the tent, and the white tunic would prevent that heat which entered the tent from heating the rider. Therefore, with a white tent, the temperature would be lower than 48.5°C , and the rate of radiant heat transferred to the rider would be less than 20.0 W .

Exercise:**Problem: Integrated Concepts**

One 30.0°C day the relative humidity is 75.0%, and that evening the temperature drops to 20.0°C , well below the dew point. (a) How many grams of water condense from each cubic meter of air? (b) How much heat transfer occurs by this condensation? (c) What temperature increase could this cause in dry air?

Exercise:**Problem: Integrated Concepts**

Large meteors sometimes strike the Earth, converting most of their kinetic energy into thermal energy. (a) What is the kinetic energy of a 10^9 kg meteor moving at 25.0 km/s? (b) If this meteor lands in a deep ocean and 80% of its kinetic energy goes into heating water, how many kilograms of water could it raise by 5.0°C ? (c) Discuss how the energy of the meteor is more likely to be deposited in the ocean and the likely effects of that energy.

Solution:

(a) 3×10^{17} J

(b) 1×10^{13} kg

(c) When a large meteor hits the ocean, it causes great tidal waves, dissipating large amount of its energy in the form of kinetic energy of the water.

Exercise:

Problem: Integrated Concepts

Frozen waste from airplane toilets has sometimes been accidentally ejected at high altitude. Ordinarily it breaks up and disperses over a large area, but sometimes it holds together and strikes the ground. Calculate the mass of 0°C ice that can be melted by the conversion of kinetic and gravitational potential energy when a 20.0 kg piece of frozen waste is released at 12.0 km altitude while moving at 250 m/s and strikes the ground at 100 m/s (since less than 20.0 kg melts, a significant mess results).

Exercise:

Problem: Integrated Concepts

(a) A large electrical power facility produces 1600 MW of “waste heat,” which is dissipated to the environment in cooling towers by warming air flowing through the towers by 5.00°C . What is the

necessary flow rate of air in m^3/s ? (b) Is your result consistent with the large cooling towers used by many large electrical power plants?

Solution:

(a) $3.44 \times 10^5 \text{ m}^3/\text{s}$

(b) This is equivalent to 12 million cubic feet of air per second. That is tremendous. This is too large to be dissipated by heating the air by only 5°C . Many of these cooling towers use the circulation of cooler air over warmer water to increase the rate of evaporation. This would allow much smaller amounts of air necessary to remove such a large amount of heat because evaporation removes larger quantities of heat than was considered in part (a).

Exercise:

Problem: Integrated Concepts

(a) Suppose you start a workout on a Stairmaster, producing power at the same rate as climbing 116 stairs per minute. Assuming your mass is 76.0 kg and your efficiency is 20.0% , how long will it take for your body temperature to rise 1.00°C if all other forms of heat transfer in and out of your body are balanced? (b) Is this consistent with your experience in getting warm while exercising?

Exercise:

Problem: Integrated Concepts

A 76.0-kg person suffering from hypothermia comes indoors and shivers vigorously. How long does it take the heat transfer to increase the person's body temperature by 2.00°C if all other forms of heat transfer are balanced?

Solution:

20.9 min

Exercise:**Problem: Integrated Concepts**

In certain large geographic regions, the underlying rock is hot. Wells can be drilled and water circulated through the rock for heat transfer for the generation of electricity. (a) Calculate the heat transfer that can be extracted by cooling 1.00 km^3 of granite by 100°C . (b) How long will this take if heat is transferred at a rate of 300 MW, assuming no heat transfers back into the 1.00 km of rock by its surroundings?

Exercise:**Problem: Integrated Concepts**

Heat transfers from your lungs and breathing passages by evaporating water. (a) Calculate the maximum number of grams of water that can be evaporated when you inhale 1.50 L of 37°C air with an original relative humidity of 40.0% . (Assume that body temperature is also 37°C .) (b) How many joules of energy are required to evaporate this amount? (c) What is the rate of heat transfer in watts from this method, if you breathe at a normal resting rate of 10.0 breaths per minute?

Solution:

(a) $3.96 \times 10^{-2} \text{ g}$

(b) 96.2 J

(c) 16.0 W

Exercise:**Problem: Integrated Concepts**

(a) What is the temperature increase of water falling 55.0 m over Niagara Falls? (b) What fraction must evaporate to keep the temperature constant?

Exercise:**Problem: Integrated Concepts**

Hot air rises because it has expanded. It then displaces a greater volume of cold air, which increases the buoyant force on it. (a) Calculate the ratio of the buoyant force to the weight of 50.0°C air surrounded by 20.0°C air. (b) What energy is needed to cause 1.00 m^3 of air to go from 20.0°C to 50.0°C ? (c) What gravitational potential energy is gained by this volume of air if it rises 1.00 m ? Will this cause a significant cooling of the air?

Solution:

(a) 1.102

(b) $2.79 \times 10^4\text{ J}$

(c) 12.6 J. This will not cause a significant cooling of the air because it is much less than the energy found in part (b), which is the energy required to warm the air from 20.0°C to 50.0°C .

Exercise:**Problem: Unreasonable Results**

(a) What is the temperature increase of an 80.0 kg person who consumes 2500 kcal of food in one day with 95.0% of the energy transferred as heat to the body? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) 36°C

(b) Any temperature increase greater than about 3°C would be unreasonably large. In this case the final temperature of the person would rise to 73°C (163°F).

(c) The assumption of 95% heat retention is unreasonable.

Exercise:

Problem: Unreasonable Results

A slightly deranged Arctic inventor surrounded by ice thinks it would be much less mechanically complex to cool a car engine by melting ice on it than by having a water-cooled system with a radiator, water pump, antifreeze, and so on. (a) If 80.0% of the energy in 1.00 gal of gasoline is converted into “waste heat” in a car engine, how many kilograms of 0°C ice could it melt? (b) Is this a reasonable amount of ice to carry around to cool the engine for 1.00 gal of gasoline consumption? (c) What premises or assumptions are unreasonable?

Exercise:

Problem: Unreasonable Results

(a) Calculate the rate of heat transfer by conduction through a window with an area of 1.00 m^2 that is 0.750 cm thick, if its inner surface is at 22.0°C and its outer surface is at 35.0°C. (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Solution:

(a) 1.46 kW

(b) Very high power loss through a window. An electric heater of this power can keep an entire room warm.

(c) The surface temperatures of the window do not differ by as great an amount as assumed. The inner surface will be warmer, and the outer surface will be cooler.

Exercise:

Problem: Unreasonable Results

A meteorite 1.20 cm in diameter is so hot immediately after penetrating the atmosphere that it radiates 20.0 kW of power. (a) What is its temperature, if the surroundings are at 20.0°C and it has an emissivity of 0.800? (b) What is unreasonable about this result? (c) Which premise or assumption is responsible?

Exercise:

Problem: Construct Your Own Problem

Consider a new model of commercial airplane having its brakes tested as a part of the initial flight permission procedure. The airplane is brought to takeoff speed and then stopped with the brakes alone. Construct a problem in which you calculate the temperature increase of the brakes during this process. You may assume most of the kinetic energy of the airplane is converted to thermal energy in the brakes and surrounding materials, and that little escapes. Note that the brakes are expected to become so hot in this procedure that they ignite and, in order to pass the test, the airplane must be able to withstand the fire for some time without a general conflagration.

Exercise:

Problem: Construct Your Own Problem

Consider a person outdoors on a cold night. Construct a problem in which you calculate the rate of heat transfer from the person by all three heat transfer methods. Make the initial circumstances such that at rest the person will have a net heat transfer and then decide how much physical activity of a chosen type is necessary to balance the rate of heat transfer. Among the things to consider are the size of the person, type of clothing, initial metabolic rate, sky conditions, amount of water evaporated, and volume of air breathed. Of course, there are many other factors to consider and your instructor may wish to guide you in the assumptions made as well as the detail of analysis and method of presenting your results.

Glossary

emissivity

measure of how well an object radiates

greenhouse effect

warming of the Earth that is due to gases such as carbon dioxide and methane that absorb infrared radiation from the Earth's surface and reradiate it in all directions, thus sending a fraction of it back toward the surface of the Earth

net rate of heat transfer by radiation

is $\frac{Q_{\text{net}}}{t} = \sigma e A (T_2^4 - T_1^4)$

radiation

energy transferred by electromagnetic waves directly as a result of a temperature difference

Stefan-Boltzmann law of radiation

$\frac{Q}{t} = \sigma e A T^4$, where σ is the Stefan-Boltzmann constant, A is the surface area of the object, T is the absolute temperature, and e is the emissivity

The First Law of Thermodynamics

- Define the first law of thermodynamics.
- Describe how conservation of energy relates to the first law of thermodynamics.
- Identify instances of the first law of thermodynamics working in everyday situations, including biological metabolism.
- Calculate changes in the internal energy of a system, after accounting for heat transfer and work done.



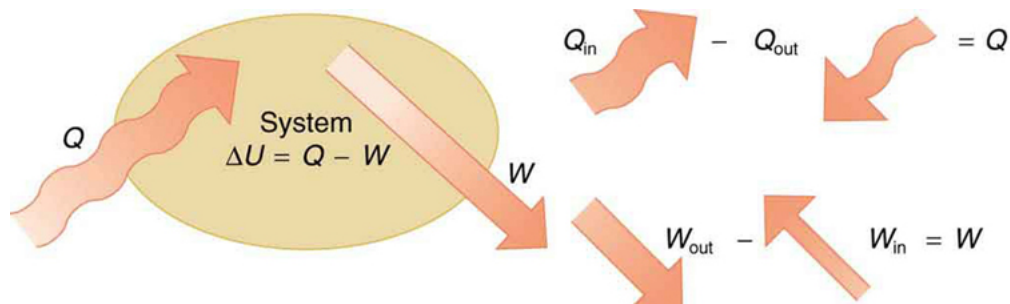
This boiling tea kettle represents energy in motion. The water in the kettle is turning to water vapor because heat is being transferred from the stove to the kettle. As the entire system gets hotter, work is done—from the evaporation of the water to the whistling of the kettle. (credit: Gina Hamilton)

If we are interested in how heat transfer is converted into doing work, then the conservation of energy principle is important. The first law of thermodynamics applies the conservation of energy principle to systems where heat transfer and doing work are the methods of transferring energy into and out of the system. The **first law of thermodynamics** states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system. In equation form, the first law of thermodynamics is

Equation:

$$\Delta U = Q - W.$$

Here ΔU is the *change in internal energy* U of the system. Q is the *net heat transferred into the system*—that is, Q is the sum of all heat transfer into and out of the system. W is the *net work done by the system*—that is, W is the sum of all work done on or by the system. We use the following sign conventions: if Q is positive, then there is a net heat transfer into the system; if W is positive, then there is net work done by the system. So positive Q adds energy to the system and positive W takes energy from the system. Thus $\Delta U = Q - W$. Note also that if more heat transfer into the system occurs than work done, the difference is stored as internal energy. Heat engines are a good example of this—heat transfer into them takes place so that they can do work. (See [\[link\]](#).) We will now examine Q , W , and ΔU further.



The first law of thermodynamics is the conservation-of-energy principle stated for a system where heat and work are the methods of transferring energy for a system in

thermal equilibrium. Q represents the net heat transfer—it is the sum of all heat transfers into and out of the system. Q is positive for net heat transfer *into* the system. W is the total work done on and by the system. W is positive when more work is done *by* the system than on it. The change in the internal energy of the system, ΔU , is related to heat and work by the first law of thermodynamics, $\Delta U = Q - W$.

Note:

Making Connections: Law of Thermodynamics and Law of Conservation of Energy

The first law of thermodynamics is actually the law of conservation of energy stated in a form most useful in thermodynamics. The first law gives the relationship between heat transfer, work done, and the change in internal energy of a system.

Heat Q and Work W

Heat transfer (Q) and doing work (W) are the two everyday means of bringing energy into or taking energy out of a system. The processes are quite different. Heat transfer, a less organized process, is driven by temperature differences. Work, a quite organized process, involves a macroscopic force exerted through a distance. Nevertheless, heat and work can produce identical results. For example, both can cause a temperature increase. Heat transfer into a system, such as when the Sun warms the air in a bicycle tire, can increase its temperature, and so can work done on the system, as when the bicyclist pumps air into the tire. Once the temperature increase has occurred, it is impossible to tell whether it was caused by heat transfer or by doing work. This uncertainty is an important point. Heat transfer and work are both energy in transit—neither is stored as such in a

system. However, both can change the internal energy U of a system. Internal energy is a form of energy completely different from either heat or work.

Internal Energy U

We can think about the internal energy of a system in two different but consistent ways. The first is the atomic and molecular view, which examines the system on the atomic and molecular scale. The **internal energy** U of a system is the sum of the kinetic and potential energies of its atoms and molecules. Recall that kinetic plus potential energy is called mechanical energy. Thus internal energy is the sum of atomic and molecular mechanical energy. Because it is impossible to keep track of all individual atoms and molecules, we must deal with averages and distributions. A second way to view the internal energy of a system is in terms of its macroscopic characteristics, which are very similar to atomic and molecular average values.

Macroscopically, we define the change in internal energy ΔU to be that given by the first law of thermodynamics:

Equation:

$$\Delta U = Q - W.$$

Many detailed experiments have verified that $\Delta U = Q - W$, where ΔU is the change in total kinetic and potential energy of all atoms and molecules in a system. It has also been determined experimentally that the internal energy U of a system depends only on the state of the system and *not how it reached that state*. More specifically, U is found to be a function of a few macroscopic quantities (pressure, volume, and temperature, for example), independent of past history such as whether there has been heat transfer or work done. This independence means that if we know the state of a system, we can calculate changes in its internal energy U from a few macroscopic variables.

Note:**Making Connections: Macroscopic and Microscopic**

In thermodynamics, we often use the macroscopic picture when making calculations of how a system behaves, while the atomic and molecular picture gives underlying explanations in terms of averages and distributions. We shall see this again in later sections of this chapter. For example, in the topic of entropy, calculations will be made using the atomic and molecular view.

To get a better idea of how to think about the internal energy of a system, let us examine a system going from State 1 to State 2. The system has internal energy U_1 in State 1, and it has internal energy U_2 in State 2, no matter how it got to either state. So the change in internal energy $\Delta U = U_2 - U_1$ is independent of what caused the change. In other words, ΔU is *independent of path*. By path, we mean the method of getting from the starting point to the ending point. Why is this independence important? Note that $\Delta U = Q - W$. Both Q and W depend on path, but ΔU does not. This path independence means that internal energy U is easier to consider than either heat transfer or work done.

Example:**Calculating Change in Internal Energy: The Same Change in U is Produced by Two Different Processes**

(a) Suppose there is heat transfer of 40.00 J to a system, while the system does 10.00 J of work. Later, there is heat transfer of 25.00 J out of the system while 4.00 J of work is done on the system. What is the net change in internal energy of the system?

(b) What is the change in internal energy of a system when a total of 150.00 J of heat transfer occurs out of (from) the system and 159.00 J of work is done on the system? (See [\[link\]](#)).

Strategy

In part (a), we must first find the net heat transfer and net work done from the given information. Then the first law of thermodynamics

$(\Delta U = Q - W)$ can be used to find the change in internal energy. In part (b), the net heat transfer and work done are given, so the equation can be used directly.

Solution for (a)

The net heat transfer is the heat transfer into the system minus the heat transfer out of the system, or

Equation:

$$Q = 40.00 \text{ J} - 25.00 \text{ J} = 15.00 \text{ J}.$$

Similarly, the total work is the work done by the system minus the work done on the system, or

Equation:

$$W = 10.00 \text{ J} - 4.00 \text{ J} = 6.00 \text{ J}.$$

Thus the change in internal energy is given by the first law of thermodynamics:

Equation:

$$\Delta U = Q - W = 15.00 \text{ J} - 6.00 \text{ J} = 9.00 \text{ J}.$$

We can also find the change in internal energy for each of the two steps. First, consider 40.00 J of heat transfer in and 10.00 J of work out, or

Equation:

$$\Delta U_1 = Q_1 - W_1 = 40.00 \text{ J} - 10.00 \text{ J} = 30.00 \text{ J}.$$

Now consider 25.00 J of heat transfer out and 4.00 J of work in, or

Equation:

$$\Delta U_2 = Q_2 - W_2 = -25.00 \text{ J} - (-4.00 \text{ J}) = -21.00 \text{ J}.$$

The total change is the sum of these two steps, or

Equation:

$$\Delta U = \Delta U_1 + \Delta U_2 = 30.00 \text{ J} + (-21.00 \text{ J}) = 9.00 \text{ J}.$$

Discussion on (a)

No matter whether you look at the overall process or break it into steps, the change in internal energy is the same.

Solution for (b)

Here the net heat transfer and total work are given directly to be

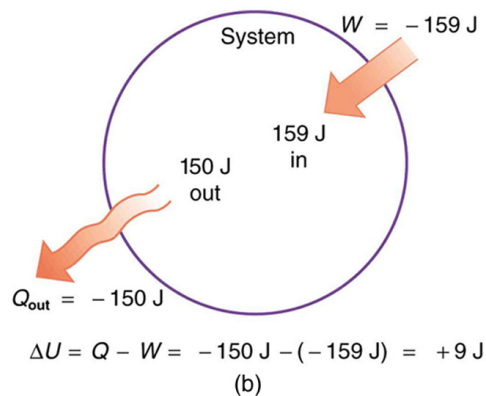
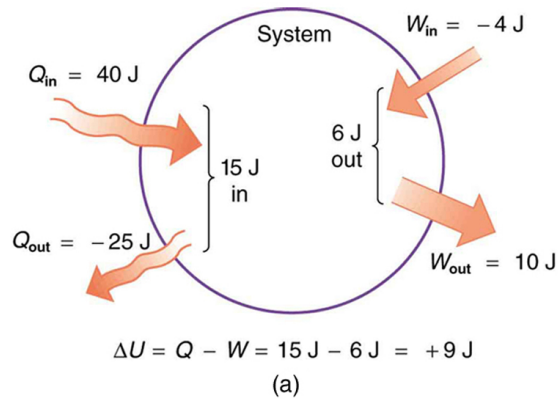
$Q = -150.00 \text{ J}$ and $W = -159.00 \text{ J}$, so that

Equation:

$$\Delta U = Q - W = -150.00 \text{ J} - (-159.00 \text{ J}) = 9.00 \text{ J}.$$

Discussion on (b)

A very different process in part (b) produces the same 9.00-J change in internal energy as in part (a). Note that the change in the system in both parts is related to ΔU and not to the individual Q s or W s involved. The system ends up in the *same* state in both (a) and (b). Parts (a) and (b) present two different paths for the system to follow between the same starting and ending points, and the change in internal energy for each is the same—it is independent of path.



Two different processes produce the same change in a system. (a) A total of 15.00 J of heat transfer occurs into the system, while work takes out a total of 6.00 J. The change in internal energy is $\Delta U = Q - W = 9.00 \text{ J}$. (b) Heat transfer removes 150.00 J from the system while work puts 159.00 J into it, producing an increase of 9.00 J in internal energy. If the system starts out in the same state in (a) and (b), it will end up in the same final state in either case—its final state is related to internal

energy, not how that energy
was acquired.

Human Metabolism and the First Law of Thermodynamics

Human metabolism is the conversion of food into heat transfer, work, and stored fat. Metabolism is an interesting example of the first law of thermodynamics in action. We now take another look at these topics via the first law of thermodynamics. Considering the body as the system of interest, we can use the first law to examine heat transfer, doing work, and internal energy in activities ranging from sleep to heavy exercise. What are some of the major characteristics of heat transfer, doing work, and energy in the body? For one, body temperature is normally kept constant by heat transfer to the surroundings. This means Q is negative. Another fact is that the body usually does work on the outside world. This means W is positive. In such situations, then, the body loses internal energy, since $\Delta U = Q - W$ is negative.

Now consider the effects of eating. Eating increases the internal energy of the body by adding chemical potential energy (this is an unromantic view of a good steak). The body *metabolizes* all the food we consume. Basically, metabolism is an oxidation process in which the chemical potential energy of food is released. This implies that food input is in the form of work. Food energy is reported in a special unit, known as the Calorie. This energy is measured by burning food in a calorimeter, which is how the units are determined.

In chemistry and biochemistry, one calorie (spelled with a *lowercase c*) is defined as the energy (or heat transfer) required to raise the temperature of one gram of pure water by one degree Celsius. Nutritionists and weight-watchers tend to use the *dietary* calorie, which is frequently called a Calorie (spelled with a *capital C*). One food Calorie is the energy needed to raise the temperature of one *kilogram* of water by one degree Celsius. This

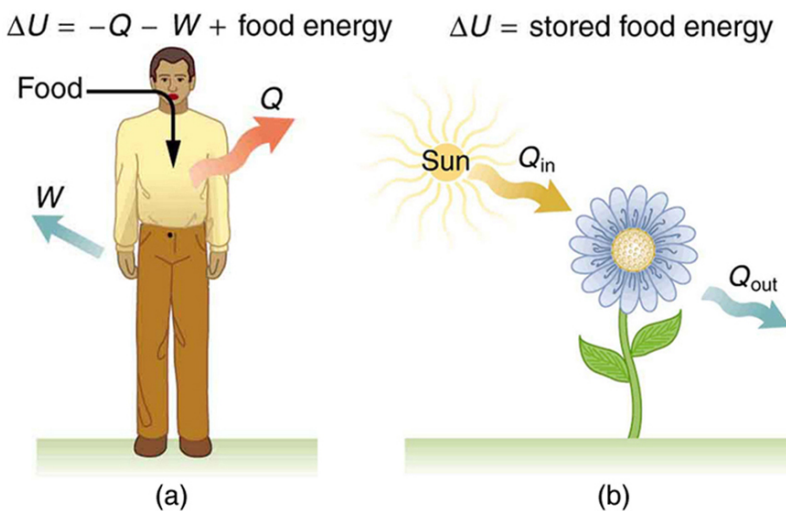
means that one dietary Calorie is equal to one kilocalorie for the chemist, and one must be careful to avoid confusion between the two.

Again, consider the internal energy the body has lost. There are three places this internal energy can go—to heat transfer, to doing work, and to stored fat (a tiny fraction also goes to cell repair and growth). Heat transfer and doing work take internal energy out of the body, and food puts it back. If you eat just the right amount of food, then your average internal energy remains constant. Whatever you lose to heat transfer and doing work is replaced by food, so that, in the long run, $\Delta U = 0$. If you overeat repeatedly, then ΔU is always positive, and your body stores this extra internal energy as fat. The reverse is true if you eat too little. If ΔU is negative for a few days, then the body metabolizes its own fat to maintain body temperature and do work that takes energy from the body. This process is how dieting produces weight loss.

Life is not always this simple, as any dieter knows. The body stores fat or metabolizes it only if energy intake changes for a period of several days. Once you have been on a major diet, the next one is less successful because your body alters the way it responds to low energy intake. Your basal metabolic rate (BMR) is the rate at which food is converted into heat transfer and work done while the body is at complete rest. The body adjusts its basal metabolic rate to partially compensate for over-eating or under-eating. The body will decrease the metabolic rate rather than eliminate its own fat to replace lost food intake. You will chill more easily and feel less energetic as a result of the lower metabolic rate, and you will not lose weight as fast as before. Exercise helps to lose weight, because it produces both heat transfer from your body and work, and raises your metabolic rate even when you are at rest. Weight loss is also aided by the quite low efficiency of the body in converting internal energy to work, so that the loss of internal energy resulting from doing work is much greater than the work done. It should be noted, however, that living systems are not in thermalequilibrium.

The body provides us with an excellent indication that many thermodynamic processes are *irreversible*. An irreversible process can go in one direction but not the reverse, under a given set of conditions. For

example, although body fat can be converted to do work and produce heat transfer, work done on the body and heat transfer into it cannot be converted to body fat. Otherwise, we could skip lunch by sunning ourselves or by walking down stairs. Another example of an irreversible thermodynamic process is photosynthesis. This process is the intake of one form of energy—light—by plants and its conversion to chemical potential energy. Both applications of the first law of thermodynamics are illustrated in [\[link\]](#). One great advantage of conservation laws such as the first law of thermodynamics is that they accurately describe the beginning and ending points of complex processes, such as metabolism and photosynthesis, without regard to the complications in between. [\[link\]](#) presents a summary of terms relevant to the first law of thermodynamics.



(a) The first law of thermodynamics applied to metabolism. Heat transferred out of the body (Q) and work done by the body (W) remove internal energy, while food intake replaces it. (Food intake may be considered as work done on the body.) (b) Plants convert part of the radiant heat transfer in sunlight to stored chemical energy, a process called photosynthesis.

Term	Definition
U	Internal energy—the sum of the kinetic and potential energies of a system's atoms and molecules. Can be divided into many subcategories, such as thermal and chemical energy. Depends only on the state of a system (such as its P , V , and T), not on how the energy entered the system. Change in internal energy is path independent.
Q	Heat—energy transferred because of a temperature difference. Characterized by random molecular motion. Highly dependent on path. Q entering a system is positive.
W	Work—energy transferred by a force moving through a distance. An organized, orderly process. Path dependent. W done by a system (either against an external force or to increase the volume of the system) is positive.

Summary of Terms for the First Law of Thermodynamics, $\Delta U = Q - W$

Section Summary

- The first law of thermodynamics is given as $\Delta U = Q - W$, where ΔU is the change in internal energy of a system, Q is the net heat transfer (the sum of all heat transfer into and out of the system), and W is the net work done (the sum of all work done on or by the system).
- Both Q and W are energy in transit; only ΔU represents an independent quantity capable of being stored.

- The internal energy U of a system depends only on the state of the system and not how it reached that state.
- Metabolism of living organisms, and photosynthesis of plants, are specialized types of heat transfer, doing work, and internal energy of systems.

Conceptual Questions

Exercise:

Problem:

Describe the photo of the tea kettle at the beginning of this section in terms of heat transfer, work done, and internal energy. How is heat being transferred? What is the work done and what is doing it? How does the kettle maintain its internal energy?

Exercise:

Problem:

The first law of thermodynamics and the conservation of energy, as discussed in [Conservation of Energy](#), are clearly related. How do they differ in the types of energy considered?

Exercise:

Problem:

Heat transfer Q and work done W are always energy in transit, whereas internal energy U is energy stored in a system. Give an example of each type of energy, and state specifically how it is either in transit or resides in a system.

Exercise:

Problem:

How do heat transfer and internal energy differ? In particular, which can be stored as such in a system and which cannot?

Exercise:

Problem:

If you run down some stairs and stop, what happens to your kinetic energy and your initial gravitational potential energy?

Exercise:**Problem:**

Give an explanation of how food energy (calories) can be viewed as molecular potential energy (consistent with the atomic and molecular definition of internal energy).

Exercise:**Problem:**

Identify the type of energy transferred to your body in each of the following as either internal energy, heat transfer, or doing work: (a) basking in sunlight; (b) eating food; (c) riding an elevator to a higher floor.

Problems & Exercises**Exercise:****Problem:**

What is the change in internal energy of a car if you put 12.0 gal of gasoline into its tank? The energy content of gasoline is $1.3 \times 10^8 \text{ J/gal}$. All other factors, such as the car's temperature, are constant.

Solution:

$$1.6 \times 10^9 \text{ J}$$

Exercise:

Problem:

How much heat transfer occurs from a system, if its internal energy decreased by 150 J while it was doing 30.0 J of work?

Exercise:**Problem:**

A system does 1.80×10^8 J of work while 7.50×10^8 J of heat transfer occurs to the environment. What is the change in internal energy of the system assuming no other changes (such as in temperature or by the addition of fuel)?

Solution:

$$-9.30 \times 10^8 \text{ J}$$

Exercise:**Problem:**

What is the change in internal energy of a system which does 4.50×10^5 J of work while 3.00×10^6 J of heat transfer occurs into the system, and 8.00×10^6 J of heat transfer occurs to the environment?

Exercise:**Problem:**

Suppose a woman does 500 J of work and 9500 J of heat transfer occurs into the environment in the process. (a) What is the decrease in her internal energy, assuming no change in temperature or consumption of food? (That is, there is no other energy transfer.) (b) What is her efficiency?

Solution:

(a) -1.0×10^4 J , or -2.39 kcal

(b) 5.00%

Exercise:**Problem:**

(a) How much food energy will a man metabolize in the process of doing 35.0 kJ of work with an efficiency of 5.00%? (b) How much heat transfer occurs to the environment to keep his temperature constant? Explicitly show how you follow the steps in the Problem-Solving Strategy for thermodynamics found in [Problem-Solving Strategies for Thermodynamics](#).

Exercise:**Problem:**

(a) What is the average metabolic rate in watts of a man who metabolizes 10,500 kJ of food energy in one day? (b) What is the maximum amount of work in joules he can do without breaking down fat, assuming a maximum efficiency of 20.0%? (c) Compare his work output with the daily output of a 187-W (0.250-horsepower) motor.

Solution:

(a) 122 W

(b) 2.10×10^6 J

(c) Work done by the motor is 1.61×10^7 J ;thus the motor produces 7.67 times the work done by the man

Exercise:**Problem:**

(a) How long will the energy in a 1470-kJ (350-kcal) cup of yogurt last in a woman doing work at the rate of 150 W with an efficiency of 20.0% (such as in leisurely climbing stairs)? (b) Does the time found in part (a) imply that it is easy to consume more food energy than you can reasonably expect to work off with exercise?

Exercise:

Problem:

(a) A woman climbing the Washington Monument metabolizes 6.00×10^2 kJ of food energy. If her efficiency is 18.0%, how much heat transfer occurs to the environment to keep her temperature constant? (b) Discuss the amount of heat transfer found in (a). Is it consistent with the fact that you quickly warm up when exercising?

Solution:

(a) 492 kJ

(b) This amount of heat is consistent with the fact that you warm quickly when exercising. Since the body is inefficient, the excess heat produced must be dissipated through sweating, breathing, etc.

Glossary

first law of thermodynamics

states that the change in internal energy of a system equals the net heat transfer *into* the system minus the net work done *by* the system

internal energy

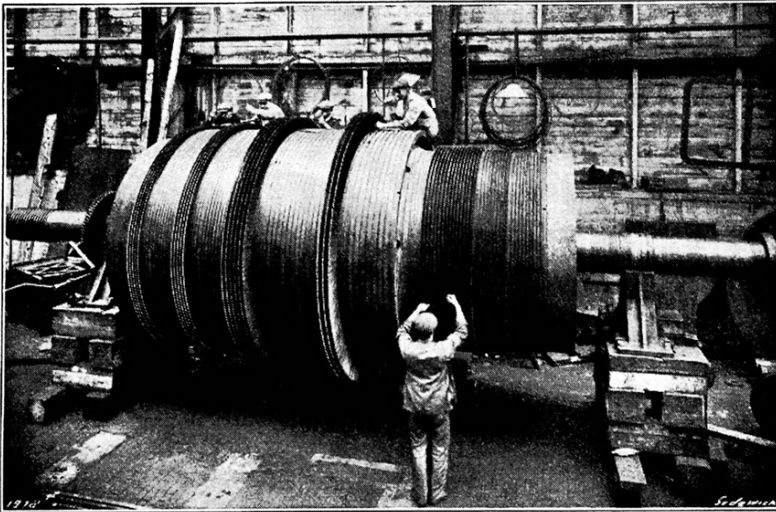
the sum of the kinetic and potential energies of a system's atoms and molecules

human metabolism

conversion of food into heat transfer, work, and stored fat

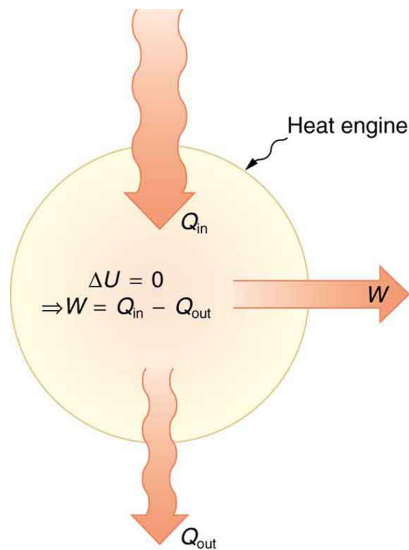
The First Law of Thermodynamics and Some Simple Processes

- Describe the processes of a simple heat engine.
- Explain the differences among the simple thermodynamic processes— isobaric, isochoric, isothermal, and adiabatic.
- Calculate total work done in a cyclical thermodynamic process.

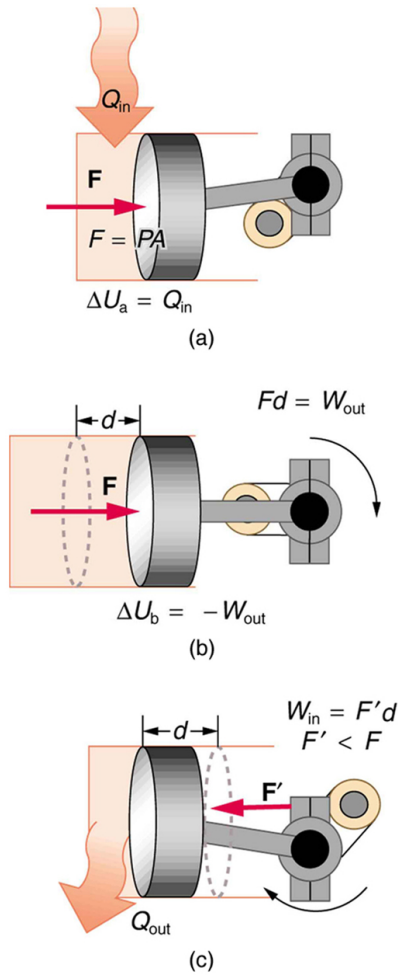


Beginning with the Industrial Revolution, humans have harnessed power through the use of the first law of thermodynamics, before we even understood it completely. This photo, of a steam engine at the Turbinia Works, dates from 1911, a mere 61 years after the first explicit statement of the first law of thermodynamics by Rudolph Clausius. (credit: public domain; author unknown)

One of the most important things we can do with heat transfer is to use it to do work for us. Such a device is called a **heat engine**. Car engines and steam turbines that generate electricity are examples of heat engines. [\[link\]](#) shows schematically how the first law of thermodynamics applies to the typical heat engine.



Schematic representation of a heat engine, governed, of course, by the first law of thermodynamics. It is impossible to devise a system where $Q_{out} = 0$, that is, in which no heat transfer occurs to the environment.



- (a) Heat transfer to the gas in a cylinder increases the internal energy of the gas, creating higher pressure and temperature. (b) The force exerted on the movable cylinder does work as the gas expands. Gas pressure and temperature decrease when it expands, indicating

that the gas's
internal energy has
been decreased by
doing work. (c)
Heat transfer to the
environment
further reduces
pressure in the gas
so that the piston
can be more easily
returned to its
starting position.

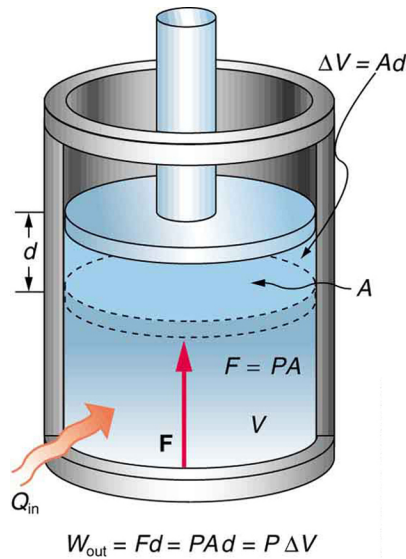
The illustrations above show one of the ways in which heat transfer does work. Fuel combustion produces heat transfer to a gas in a cylinder, increasing the pressure of the gas and thereby the force it exerts on a movable piston. The gas does work on the outside world, as this force moves the piston through some distance. Heat transfer to the gas cylinder results in work being done. To repeat this process, the piston needs to be returned to its starting point. Heat transfer now occurs from the gas to the surroundings so that its pressure decreases, and a force is exerted by the surroundings to push the piston back through some distance. Variations of this process are employed daily in hundreds of millions of heat engines. We will examine heat engines in detail in the next section. In this section, we consider some of the simpler underlying processes on which heat engines are based.

PV Diagrams and their Relationship to Work Done on or by a Gas

A process by which a gas does work on a piston at constant pressure is called an **isobaric process**. Since the pressure is constant, the force exerted is constant and the work done is given as

Equation:

$$P\Delta V.$$



An isobaric expansion of a gas requires heat transfer to keep the pressure constant. Since pressure is constant, the work done is $P\Delta V$.

Equation:

$$W = Fd$$

See the symbols as shown in [\[link\]](#). Now $F = PA$, and so

Equation:

$$W = PA d.$$

Because the volume of a cylinder is its cross-sectional area A times its length d , we see that $Ad = \Delta V$, the change in volume; thus,

Equation:

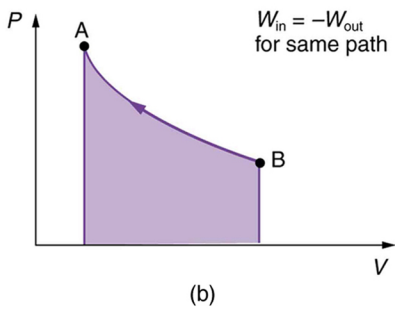
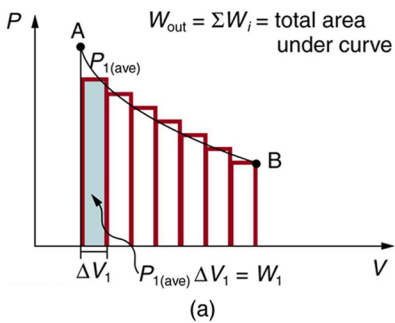
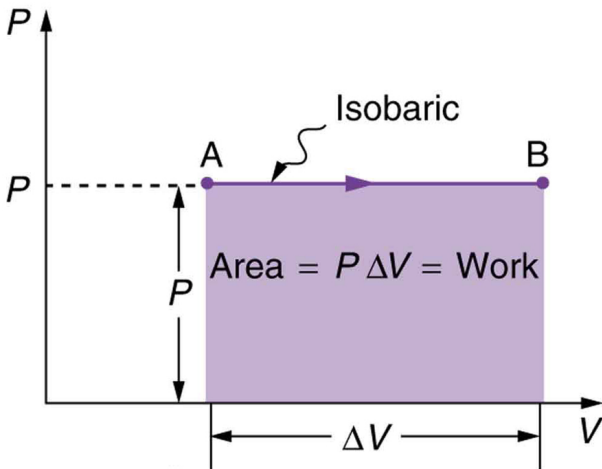
$$W = P\Delta V \text{ (isobaric process).}$$

Note that if ΔV is positive, then W is positive, meaning that work is done by the gas on the outside world.

(Note that the pressure involved in this work that we've called P is the pressure of the gas *inside* the tank. If we call the pressure outside the tank P_{ext} , an expanding gas would be working *against* the external pressure; the work done would therefore be $W = -P_{\text{ext}}\Delta V$ (isobaric process). Many texts use this definition of work, and not the definition based on internal pressure, as the basis of the First Law of Thermodynamics. This definition reverses the sign conventions for work, and results in a statement of the first law that becomes $\Delta U = Q + W$.)

It is not surprising that $W = P\Delta V$, since we have already noted in our treatment of fluids that pressure is a type of potential energy per unit volume and that pressure in fact has units of energy divided by volume. We also noted in our discussion of the ideal gas law that PV has units of energy. In this case, some of the energy associated with pressure becomes work.

[\[link\]](#) shows a graph of pressure versus volume (that is, a PV diagram for an isobaric process. You can see in the figure that the work done is the area under the graph. This property of PV diagrams is very useful and broadly applicable: *the work done on or by a system in going from one state to another equals the area under the curve on a PV diagram.*

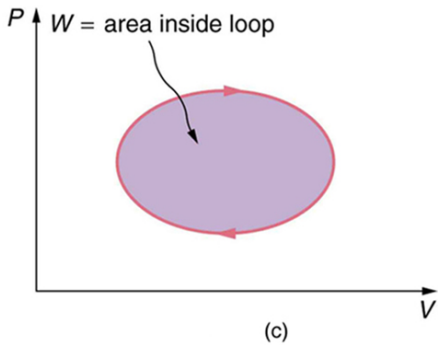
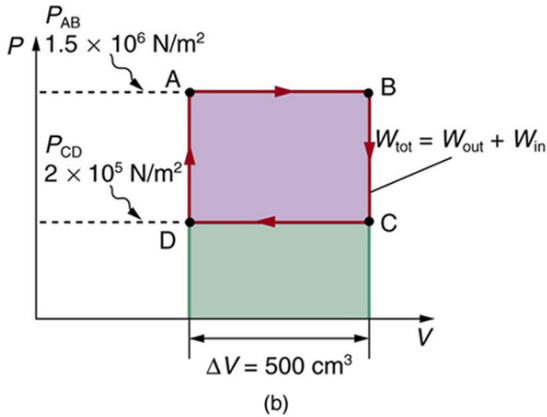
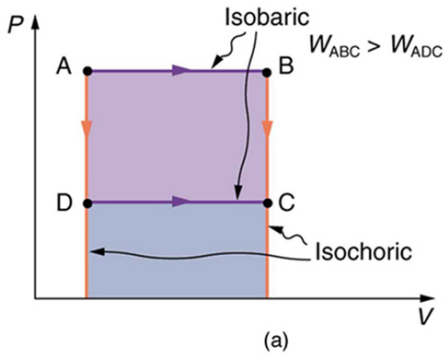


(a) A PV diagram in which pressure varies as well as volume. The work done for each interval is its average pressure times the change in volume, or the area under the curve over that interval. Thus the total area under the curve equals the total work done. (b) Work must be done on the system to follow the reverse path. This is interpreted as a negative area under the curve.

We can see where this leads by considering [\[link\]](#)(a), which shows a more general process in which both pressure and volume change. The area under the curve is closely approximated by dividing it into strips, each having an average constant pressure $P_{i(\text{ave})}$. The work done is $W_i = P_{i(\text{ave})}\Delta V_i$ for each strip, and the total work done is the sum of the W_i . Thus the total work done is the total area under the curve. If the path is reversed, as in [\[link\]](#)(b), then work is done on the system. The area under the curve in that case is negative, because ΔV is negative.

PV diagrams clearly illustrate that *the work done depends on the path taken and not just the endpoints*. This path dependence is seen in [\[link\]](#)(a), where

more work is done in going from A to C by the path via point B than by the path via point D. The vertical paths, where volume is constant, are called **isochoric** processes. Since volume is constant, $\Delta V = 0$, and no work is done in an isochoric process. Now, if the system follows the cyclical path ABCDA, as in [\[link\]](#)(b), then the total work done is the area inside the loop. The negative area below path CD subtracts, leaving only the area inside the rectangle. In fact, the work done in any cyclical process (one that returns to its starting point) is the area inside the loop it forms on a PV diagram, as [\[link\]](#)(c) illustrates for a general cyclical process. Note that the loop must be traversed in the clockwise direction for work to be positive—that is, for there to be a net work output.



(a) The work done in going from A to C depends on path. The work is greater for the path ABC than for the path ADC, because the former is at higher pressure. In both cases, the work done is the area under the path. This area is greater for path ABC. (b) The total work done in the cyclical process

ABCD is the area inside the loop, since the negative area below CD subtracts out, leaving just the area inside the rectangle. (The values given for the pressures and the change in volume are intended for use in the example below.) (c) The area inside any closed loop is the work done in the cyclical process. If the loop is traversed in a clockwise direction, W is positive—it is work done on the outside environment. If the loop is traveled in a counter-clockwise direction, W is negative—it is work that is done to the system.

Example:**Total Work Done in a Cyclical Process Equals the Area Inside the Closed Loop on a PV Diagram**

Calculate the total work done in the cyclical process ABCDA shown in [\[link\]](#)(b) by the following two methods to verify that work equals the area inside the closed loop on the PV diagram. (Take the data in the figure to be precise to three significant figures.) (a) Calculate the work done along each segment of the path and add these values to get the total work. (b) Calculate the area inside the rectangle ABCDA.

Strategy

To find the work along any path on a PV diagram, you use the fact that work is pressure times change in volume, or $W = P\Delta V$. So in part (a),

this value is calculated for each leg of the path around the closed loop.

Solution for (a)

The work along path AB is

Equation:

$$\begin{aligned}W_{AB} &= P_{AB}\Delta V_{AB} \\&= (1.50\times 10^6 \text{ N/m}^2)(5.00\times 10^{-4} \text{ m}^3) = 750 \text{ J}.\end{aligned}$$

Since the path BC is isochoric, $\Delta V_{BC} = 0$, and so $W_{BC} = 0$. The work along path CD is negative, since ΔV_{CD} is negative (the volume decreases). The work is

Equation:

$$\begin{aligned}W_{CD} &= P_{CD}\Delta V_{CD} \\&= (2.00\times 10^5 \text{ N/m}^2)(-5.00\times 10^{-4} \text{ m}^3) = -100 \text{ J}.\end{aligned}$$

Again, since the path DA is isochoric, $\Delta V_{DA} = 0$, and so $W_{DA} = 0$. Now the total work is

Equation:

$$\begin{aligned}W &= W_{AB} + W_{BC} + W_{CD} + W_{DA} \\&= 750 \text{ J} + 0 + (-100\text{J}) + 0 = 650 \text{ J}.\end{aligned}$$

Solution for (b)

The area inside the rectangle is its height times its width, or

Equation:

$$\begin{aligned}\text{area} &= (P_{AB} - P_{CD})\Delta V \\&= \left[(1.50\times 10^6 \text{ N/m}^2) - (2.00\times 10^5 \text{ N/m}^2) \right] (5.00\times 10^{-4} \text{ m}^3) \\&= 650 \text{ J}.\end{aligned}$$

Thus,

Equation:

$$\text{area} = 650 \text{ J} = W.$$

Discussion

The result, as anticipated, is that the area inside the closed loop equals the work done. The area is often easier to calculate than is the work done along each path. It is also convenient to visualize the area inside different curves on PV diagrams in order to see which processes might produce the most work. Recall that work can be done to the system, or by the system, depending on the sign of W . A positive W is work that is done by the system on the outside environment; a negative W represents work done by the environment on the system.

[\[link\]](#)(a) shows two other important processes on a PV diagram. For comparison, both are shown starting from the same point A. The upper curve ending at point B is an **isothermal** process—that is, one in which temperature is kept constant. If the gas behaves like an ideal gas, as is often the case, and if no phase change occurs, then $PV = nRT$. Since T is constant, PV is a constant for an isothermal process. We ordinarily expect the temperature of a gas to decrease as it expands, and so we correctly suspect that heat transfer must occur from the surroundings to the gas to keep the temperature constant during an isothermal expansion. To show this more rigorously for the special case of a monatomic ideal gas, we note that the average kinetic energy of an atom in such a gas is given by

Equation:

$$\frac{1}{2}mv^2 = \frac{3}{2}kT.$$

The kinetic energy of the atoms in a monatomic ideal gas is its only form of internal energy, and so its total internal energy U is

Equation:

$$U = N\frac{1}{2}mv^2 = \frac{3}{2}NkT, \text{ (monatomic ideal gas),}$$

where N is the number of atoms in the gas. This relationship means that the internal energy of an ideal monatomic gas is constant during an isothermal process—that is, $\Delta U = 0$. If the internal energy does not change, then the net heat transfer into the gas must equal the net work done by the gas. That is, because $\Delta U = Q - W = 0$ here, $Q = W$. We must have just enough heat transfer to replace the work done. An isothermal

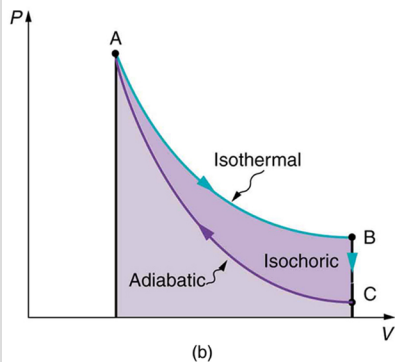
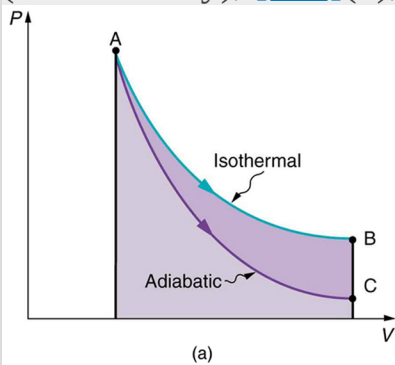
process is inherently slow, because heat transfer occurs continuously to keep the gas temperature constant at all times and must be allowed to spread through the gas so that there are no hot or cold regions.

Also shown in [\[link\]](#)(a) is a curve AC for an **adiabatic** process, defined to be one in which there is no heat transfer—that is, $Q = 0$. Processes that are nearly adiabatic can be achieved either by using very effective insulation or by performing the process so fast that there is little time for heat transfer. Temperature must decrease during an adiabatic expansion process, since work is done at the expense of internal energy:

Equation:

$$U = \frac{3}{2}NkT.$$

(You might have noted that a gas released into atmospheric pressure from a pressurized cylinder is substantially colder than the gas in the cylinder.) In fact, because $Q = 0$, $\Delta U = -W$ for an adiabatic process. Lower temperature results in lower pressure along the way, so that curve AC is lower than curve AB, and less work is done. If the path ABCA could be followed by cooling the gas from B to C at constant volume (isochorically), [\[link\]](#)(b), there would be a net work output.



(a) The upper curve is an isothermal process ($\Delta T = 0$), whereas the lower curve is an adiabatic process ($Q = 0$). Both start from the same point A, but the isothermal process does more work than the adiabatic because heat transfer into the gas takes place to keep its temperature constant. This keeps the pressure higher all along the isothermal path than along the adiabatic path, producing more work. The adiabatic path thus ends up with a lower pressure and temperature at point C, even though the final volume is the same as for the isothermal process.

(b) The cycle ABCA produces a net work output.

Reversible Processes

Both isothermal and adiabatic processes such as shown in [\[link\]](#) are reversible in principle. A **reversible process** is one in which both the system and its environment can return to exactly the states they were in by following the reverse path. The reverse isothermal and adiabatic paths are BA and CA, respectively. Real macroscopic processes are never exactly reversible. In the previous examples, our system is a gas (like that in [\[link\]](#)), and its environment is the piston, cylinder, and the rest of the universe. If there are any energy-dissipating mechanisms, such as friction or turbulence, then heat transfer to the environment occurs for either direction of the piston. So, for example, if the path BA is followed and there is friction, then the gas will be returned to its original state but the environment will not—it will have been heated in both directions. Reversibility requires the direction of heat transfer to reverse for the reverse path. Since dissipative mechanisms cannot be completely eliminated, real processes cannot be reversible.

There must be reasons that real macroscopic processes cannot be reversible. We can imagine them going in reverse. For example, heat transfer occurs spontaneously from hot to cold and never spontaneously the reverse. Yet it would not violate the first law of thermodynamics for this to happen. In fact, all spontaneous processes, such as bubbles bursting, never go in reverse. There is a second thermodynamic law that forbids them from going in reverse. When we study this law, we will learn something about nature and also find that such a law limits the efficiency of heat engines. We will find that heat engines with the greatest possible theoretical efficiency would have to use reversible processes, and even they cannot convert all heat transfer into doing work. [\[link\]](#) summarizes the simpler thermodynamic processes and their definitions.

Isobaric	Constant pressure $W = P\Delta V$
Isochoric	Constant volume $W = 0$
Isothermal	Constant temperature $Q = W$
Adiabatic	No heat transfer $Q = 0$

Summary of Simple Thermodynamic Processes

Note:

PhET Explorations: States of Matter

Watch different types of molecules form a solid, liquid, or gas. Add or remove heat and watch the phase change. Change the temperature or volume of a container and see a pressure-temperature diagram respond in real time. Relate the interaction potential to the forces between molecules.

https://phet.colorado.edu/sims/html/states-of-matter/latest/states-of-matter_en.html

Section Summary

- One of the important implications of the first law of thermodynamics is that machines can be harnessed to do work that humans previously

did by hand or by external energy supplies such as running water or the heat of the Sun. A machine that uses heat transfer to do work is known as a heat engine.

- There are several simple processes, used by heat engines, that flow from the first law of thermodynamics. Among them are the isobaric, isochoric, isothermal and adiabatic processes.
- These processes differ from one another based on how they affect pressure, volume, temperature, and heat transfer.
- If the work done is performed on the outside environment, work (W) will be a positive value. If the work done is done to the heat engine system, work (W) will be a negative value.
- Some thermodynamic processes, including isothermal and adiabatic processes, are reversible in theory; that is, both the thermodynamic system and the environment can be returned to their initial states. However, because of loss of energy owing to the second law of thermodynamics, complete reversibility does not work in practice.

Conceptual Questions

Exercise:

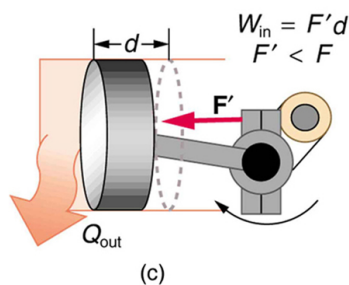
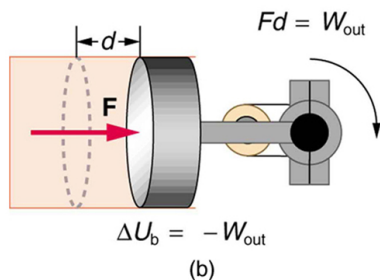
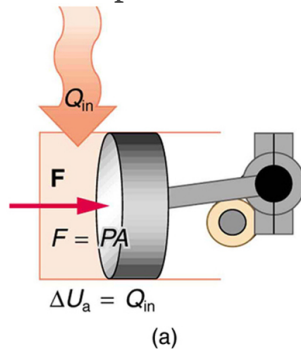
Problem:

A great deal of effort, time, and money has been spent in the quest for the so-called perpetual-motion machine, which is defined as a hypothetical machine that operates or produces useful work indefinitely and/or a hypothetical machine that produces more work or energy than it consumes. Explain, in terms of heat engines and the first law of thermodynamics, why or why not such a machine is likely to be constructed.

Exercise:

Problem:

One method of converting heat transfer into doing work is for heat transfer into a gas to take place, which expands, doing work on a piston, as shown in the figure below. (a) Is the heat transfer converted directly to work in an isobaric process, or does it go through another form first? Explain your answer. (b) What about in an isothermal process? (c) What about in an adiabatic process (where heat transfer occurred prior to the adiabatic process)?



Exercise:

Problem:

Would the previous question make any sense for an isochoric process? Explain your answer.

Exercise:**Problem:**

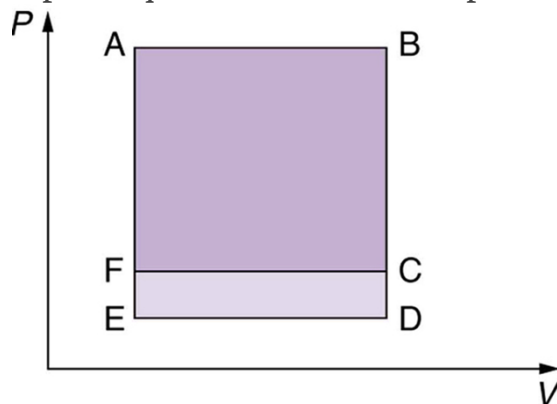
We ordinarily say that $\Delta U = 0$ for an isothermal process. Does this assume no phase change takes place? Explain your answer.

Exercise:**Problem:**

The temperature of a rapidly expanding gas decreases. Explain why in terms of the first law of thermodynamics. (Hint: Consider whether the gas does work and whether heat transfer occurs rapidly into the gas through conduction.)

Exercise:**Problem:**

Which cyclical process represented by the two closed loops, ABCFA and ABDEA, on the PV diagram in the figure below produces the greatest *net* work? Is that process also the one with the smallest work input required to return it to point A? Explain your responses.



The two cyclical processes shown on this PV diagram start with and return the system to the conditions at point A, but they follow

different paths and produce different amounts of work.

Exercise:

Problem:

A real process may be nearly adiabatic if it occurs over a very short time. How does the short time span help the process to be adiabatic?

Exercise:

Problem:

It is unlikely that a process can be isothermal unless it is a very slow process. Explain why. Is the same true for isobaric and isochoric processes? Explain your answer.

Problem Exercises

Exercise:

Problem:

A car tire contains 0.0380 m^3 of air at a pressure of $2.20 \times 10^5 \text{ N/m}^2$ (about 32 psi). How much more internal energy does this gas have than the same volume has at zero gauge pressure (which is equivalent to normal atmospheric pressure)?

Solution:

$$6.77 \times 10^3 \text{ J}$$

Exercise:

Problem:

A helium-filled toy balloon has a gauge pressure of 0.200 atm and a volume of 10.0 L. How much greater is the internal energy of the helium in the balloon than it would be at zero gauge pressure?

Exercise:**Problem:**

Steam to drive an old-fashioned steam locomotive is supplied at a constant gauge pressure of $1.75 \times 10^6 \text{ N/m}^2$ (about 250 psi) to a piston with a 0.200-m radius. (a) By calculating $P\Delta V$, find the work done by the steam when the piston moves 0.800 m. Note that this is the net work output, since gauge pressure is used. (b) Now find the amount of work by calculating the force exerted times the distance traveled. Is the answer the same as in part (a)?

Solution:

(a) $W = P\Delta V = 1.76 \times 10^5 \text{ J}$

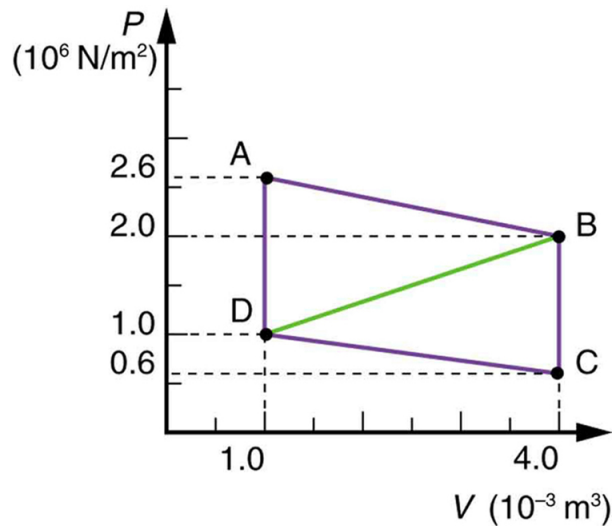
(b) $W = Fd = 1.76 \times 10^5 \text{ J}$. Yes, the answer is the same.

Exercise:**Problem:**

A hand-driven tire pump has a piston with a 2.50-cm diameter and a maximum stroke of 30.0 cm. (a) How much work do you do in one stroke if the average gauge pressure is $2.40 \times 10^5 \text{ N/m}^2$ (about 35 psi)? (b) What average force do you exert on the piston, neglecting friction and gravitational force?

Exercise:**Problem:**

Calculate the net work output of a heat engine following path ABCDA in the figure below.



Solution:

$$W = 4.5 \times 10^3 \text{ J}$$

Exercise:

Problem:

What is the net work output of a heat engine that follows path ABDA in the figure above, with a straight line from B to D? Why is the work output less than for path ABCDA? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Thermodynamics](#).

Exercise:

Problem: Unreasonable Results

What is wrong with the claim that a cyclical heat engine does 4.00 kJ of work on an input of 24.0 kJ of heat transfer while 16.0 kJ of heat transfers to the environment?

Solution:

W is not equal to the difference between the heat input and the heat output.

Exercise:

Problem:

- (a) A cyclical heat engine, operating between temperatures of 450°C and 150°C produces 4.00 MJ of work on a heat transfer of 5.00 MJ into the engine. How much heat transfer occurs to the environment?
(b) What is unreasonable about the engine? (c) Which premise is unreasonable?

Exercise:**Problem: Construct Your Own Problem**

Consider a car's gasoline engine. Construct a problem in which you calculate the maximum efficiency this engine can have. Among the things to consider are the effective hot and cold reservoir temperatures. Compare your calculated efficiency with the actual efficiency of car engines.

Exercise:**Problem: Construct Your Own Problem**

Consider a car trip into the mountains. Construct a problem in which you calculate the overall efficiency of the car for the trip as a ratio of kinetic and potential energy gained to fuel consumed. Compare this efficiency to the thermodynamic efficiency quoted for gasoline engines and discuss why the thermodynamic efficiency is so much greater. Among the factors to be considered are the gain in altitude and speed, the mass of the car, the distance traveled, and typical fuel economy.

Glossary

heat engine

a machine that uses heat transfer to do work

isobaric process

constant-pressure process in which a gas does work

isochoric process

a constant-volume process

isothermal process

a constant-temperature process

adiabatic process

a process in which no heat transfer takes place

reversible process

a process in which both the heat engine system and the external environment theoretically can be returned to their original states

Introduction to the Second Law of Thermodynamics: Heat Engines and Their Efficiency

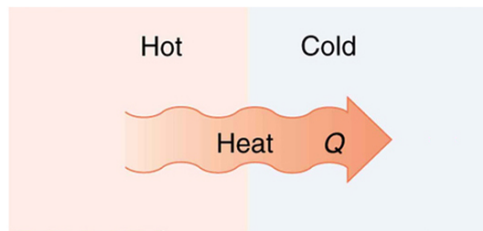
- State the expressions of the second law of thermodynamics.
- Calculate the efficiency and carbon dioxide emission of a coal-fired electricity plant, using second law characteristics.
- Describe and define the Otto cycle.



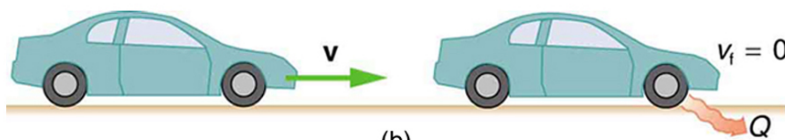
These ice floes melt during the Arctic summer. Some of them refreeze in the winter, but the second law of thermodynamics predicts that it would be extremely unlikely for the water molecules contained in these particular floes to reform the distinctive alligator-like shape they formed when the picture was taken in the summer of 2009. (credit: Patrick Kelley, U.S. Coast Guard, U.S. Geological Survey)

The second law of thermodynamics deals with the direction taken by spontaneous processes. Many processes occur spontaneously in one direction only—that is, they are irreversible, under a given set of

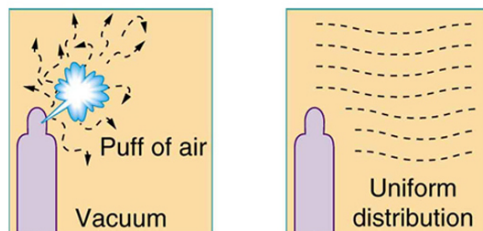
conditions. Although irreversibility is seen in day-to-day life—a broken glass does not resume its original state, for instance—complete irreversibility is a statistical statement that cannot be seen during the lifetime of the universe. More precisely, an **irreversible process** is one that depends on path. If the process can go in only one direction, then the reverse path differs fundamentally and the process cannot be reversible. For example, as noted in the previous section, heat involves the transfer of energy from higher to lower temperature. A cold object in contact with a hot one never gets colder, transferring heat to the hot object and making it hotter. Furthermore, mechanical energy, such as kinetic energy, can be completely converted to thermal energy by friction, but the reverse is impossible. A hot stationary object never spontaneously cools off and starts moving. Yet another example is the expansion of a puff of gas introduced into one corner of a vacuum chamber. The gas expands to fill the chamber, but it never regroups in the corner. The random motion of the gas molecules could take them all back to the corner, but this is never observed to happen. (See [\[link\]](#).)



(a)



(b)



(c)

Examples of one-way processes in nature.

(a) Heat transfer occurs spontaneously from hot to cold and not from cold to hot. (b) The brakes of this car convert its kinetic energy to heat transfer to the environment. The reverse process is impossible. (c) The burst of gas let into this vacuum chamber quickly expands to uniformly fill every part of the chamber. The random motions of the gas molecules will never return them to the corner.

The fact that certain processes never occur suggests that there is a law forbidding them to occur. The first law of thermodynamics would allow them to occur—none of those processes violate conservation of energy. The law that forbids these processes is called the second law of thermodynamics. We shall see that the second law can be stated in many ways that may seem different, but which in fact are equivalent. Like all natural laws, the second law of thermodynamics gives insights into nature, and its several statements imply that it is broadly applicable, fundamentally affecting many apparently disparate processes.

The already familiar direction of heat transfer from hot to cold is the basis of our first version of the **second law of thermodynamics**.

Note:

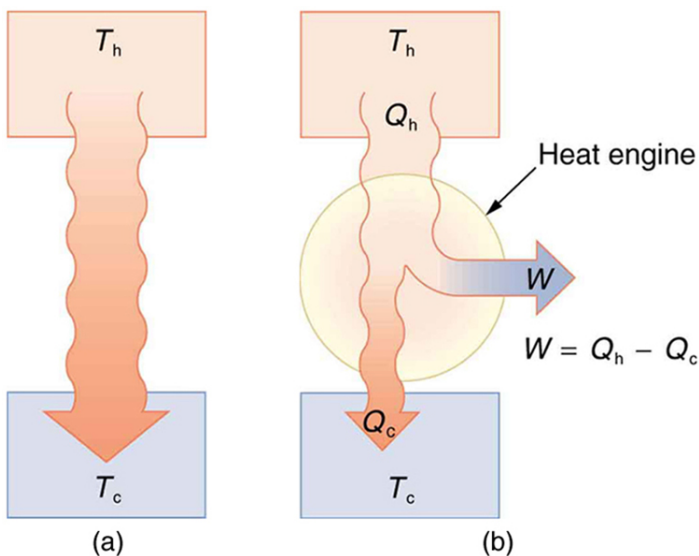
The Second Law of Thermodynamics (first expression)

Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction.

Another way of stating this: It is impossible for any process to have as its sole result heat transfer from a cooler to a hotter object.

Heat Engines

Now let us consider a device that uses heat transfer to do work. As noted in the previous section, such a device is called a heat engine, and one is shown schematically in [\[link\]](#)(b). Gasoline and diesel engines, jet engines, and steam turbines are all heat engines that do work by using part of the heat transfer from some source. Heat transfer from the hot object (or hot reservoir) is denoted as Q_h , while heat transfer into the cold object (or cold reservoir) is Q_c , and the work done by the engine is W . The temperatures of the hot and cold reservoirs are T_h and T_c , respectively.



- (a) Heat transfer occurs spontaneously from a hot object to a cold one, consistent with the second law of thermodynamics. (b) A heat engine, represented here by a circle, uses part of the heat transfer to do work. The hot and cold objects are called the hot and cold reservoirs. Q_h is the heat transfer out of the hot reservoir, W is the work output, and Q_c is the heat transfer into the cold reservoir.

Because the hot reservoir is heated externally, which is energy intensive, it is important that the work is done as efficiently as possible. In fact, we would like W to equal Q_h , and for there to be no heat transfer to the environment ($Q_c = 0$). Unfortunately, this is impossible. The **second law of thermodynamics** also states, with regard to using heat transfer to do work (the second expression of the second law):

Note:

The Second Law of Thermodynamics (second expression)

It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.

A **cyclical process** brings a system, such as the gas in a cylinder, back to its original state at the end of every cycle. Most heat engines, such as reciprocating piston engines and rotating turbines, use cyclical processes. The second law, just stated in its second form, clearly states that such engines cannot have perfect conversion of heat transfer into work done. Before going into the underlying reasons for the limits on converting heat transfer into work, we need to explore the relationships among W , Q_h , and Q_c , and to define the efficiency of a cyclical heat engine. As noted, a cyclical process brings the system back to its original condition at the end of every cycle. Such a system's internal energy U is the same at the beginning and end of every cycle—that is, $\Delta U = 0$. The first law of thermodynamics states that

Equation:

$$\Delta U = Q - W,$$

where Q is the *net* heat transfer during the cycle ($Q = Q_h - Q_c$) and W is the net work done by the system. Since $\Delta U = 0$ for a complete cycle, we

have

Equation:

$$0 = Q - W,$$

so that

Equation:

$$W = Q.$$

Thus the net work done by the system equals the net heat transfer into the system, or

Equation:

$$W = Q_h - Q_c \text{ (cyclical process),}$$

just as shown schematically in [\[link\]](#)(b). The problem is that in all processes, there is some heat transfer Q_c to the environment—and usually a very significant amount at that.

In the conversion of energy to work, we are always faced with the problem of getting less out than we put in. We define *conversion efficiency* Eff to be the ratio of useful work output to the energy input (or, in other words, the ratio of what we get to what we spend). In that spirit, we define the efficiency of a heat engine to be its net work output W divided by heat transfer to the engine Q_h ; that is,

Equation:

$$\text{Eff} = \frac{W}{Q_h}.$$

Since $W = Q_h - Q_c$ in a cyclical process, we can also express this as

Equation:

$$\text{Eff} = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h} \text{ (cyclical process),}$$

making it clear that an efficiency of 1, or 100%, is possible only if there is no heat transfer to the environment ($Q_c = 0$). Note that all Q s are positive. The direction of heat transfer is indicated by a plus or minus sign. For example, Q_c is out of the system and so is preceded by a minus sign.

Example:

Daily Work Done by a Coal-Fired Power Station, Its Efficiency and Carbon Dioxide Emissions

A coal-fired power station is a huge heat engine. It uses heat transfer from burning coal to do work to turn turbines, which are used to generate electricity. In a single day, a large coal power station has 2.50×10^{14} J of heat transfer from coal and 1.48×10^{14} J of heat transfer into the environment. (a) What is the work done by the power station? (b) What is the efficiency of the power station? (c) In the combustion process, the following chemical reaction occurs: $C + O_2 \rightarrow CO_2$. This implies that every 12 kg of coal puts $12 \text{ kg} + 16 \text{ kg} + 16 \text{ kg} = 44 \text{ kg}$ of carbon dioxide into the atmosphere. Assuming that 1 kg of coal can provide 2.5×10^6 J of heat transfer upon combustion, how much CO_2 is emitted per day by this power plant?

Strategy for (a)

We can use $W = Q_h - Q_c$ to find the work output W , assuming a cyclical process is used in the power station. In this process, water is boiled under pressure to form high-temperature steam, which is used to run steam turbine-generators, and then condensed back to water to start the cycle again.

Solution for (a)

Work output is given by:

Equation:

$$W = Q_h - Q_c.$$

Substituting the given values:

Equation:

$$\begin{aligned} W &= 2.50 \times 10^{14} \text{ J} - 1.48 \times 10^{14} \text{ J} \\ &= 1.02 \times 10^{14} \text{ J}. \end{aligned}$$

Strategy for (b)

The efficiency can be calculated with $\text{Eff} = \frac{W}{Q_h}$ since Q_h is given and work W was found in the first part of this example.

Solution for (b)

Efficiency is given by: $\text{Eff} = \frac{W}{Q_h}$. The work W was just found to be $1.02 \times 10^{14} \text{ J}$, and Q_h is given, so the efficiency is

Equation:

$$\begin{aligned} \text{Eff} &= \frac{1.02 \times 10^{14} \text{ J}}{2.50 \times 10^{14} \text{ J}} \\ &= 0.408, \text{ or } 40.8\% \end{aligned}$$

Strategy for (c)

The daily consumption of coal is calculated using the information that each day there is $2.50 \times 10^{14} \text{ J}$ of heat transfer from coal. In the combustion process, we have $\text{C} + \text{O}_2 \rightarrow \text{CO}_2$. So every 12 kg of coal puts 12 kg + 16 kg + 16 kg = 44 kg of CO_2 into the atmosphere.

Solution for (c)

The daily coal consumption is

Equation:

$$\frac{2.50 \times 10^{14} \text{ J}}{2.50 \times 10^6 \text{ J/kg}} = 1.0 \times 10^8 \text{ kg}.$$

Assuming that the coal is pure and that all the coal goes toward producing carbon dioxide, the carbon dioxide produced per day is

Equation:

$$1.0 \times 10^8 \text{ kg coal} \times \frac{44 \text{ kg CO}_2}{12 \text{ kg coal}} = 3.7 \times 10^8 \text{ kg CO}_2.$$

This is 370,000 metric tons of CO_2 produced every day.

Discussion

If all the work output is converted to electricity in a period of one day, the average power output is 1180 MW (this is left to you as an end-of-chapter problem). This value is about the size of a large-scale conventional power plant. The efficiency found is acceptably close to the value of 42% given for coal power stations. It means that fully 59.2% of the energy is heat transfer to the environment, which usually results in warming lakes, rivers, or the ocean near the power station, and is implicated in a warming planet generally. While the laws of thermodynamics limit the efficiency of such plants—including plants fired by nuclear fuel, oil, and natural gas—the heat transfer to the environment could be, and sometimes is, used for heating homes or for industrial processes. The generally low cost of energy has not made it economical to make better use of the waste heat transfer from most heat engines. Coal-fired power plants produce the greatest amount of CO_2 per unit energy output (compared to natural gas or oil), making coal the least efficient fossil fuel.

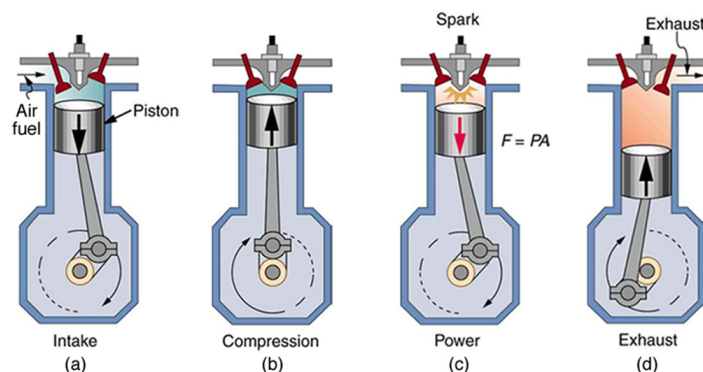
With the information given in [\[link\]](#), we can find characteristics such as the efficiency of a heat engine without any knowledge of how the heat engine operates, but looking further into the mechanism of the engine will give us greater insight. [\[link\]](#) illustrates the operation of the common four-stroke gasoline engine. The four steps shown complete this heat engine's cycle, bringing the gasoline-air mixture back to its original condition.

The **Otto cycle** shown in [\[link\]](#)(a) is used in four-stroke internal combustion engines, although in fact the true Otto cycle paths do not correspond exactly to the strokes of the engine.

The adiabatic process AB corresponds to the nearly adiabatic compression stroke of the gasoline engine. In both cases, work is done on the system (the gas mixture in the cylinder), increasing its temperature and pressure. Along path BC of the Otto cycle, heat transfer Q_h into the gas occurs at constant volume, causing a further increase in pressure and temperature. This process corresponds to burning fuel in an internal combustion engine, and takes place so rapidly that the volume is nearly constant. Path CD in the Otto cycle is an adiabatic expansion that does work on the outside world,

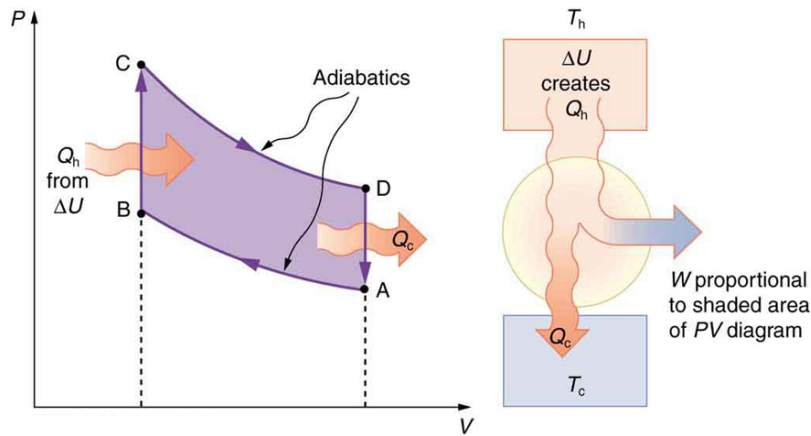
just as the power stroke of an internal combustion engine does in its nearly adiabatic expansion. The work done by the system along path CD is greater than the work done on the system along path AB, because the pressure is greater, and so there is a net work output. Along path DA in the Otto cycle, heat transfer Q_c from the gas at constant volume reduces its temperature and pressure, returning it to its original state. In an internal combustion engine, this process corresponds to the exhaust of hot gases and the intake of an air-gasoline mixture at a considerably lower temperature. In both cases, heat transfer into the environment occurs along this final path.

The net work done by a cyclical process is the area inside the closed path on a PV diagram, such as that inside path ABCDA in [\[link\]](#). Note that in every imaginable cyclical process, it is absolutely necessary for heat transfer from the system to occur in order to get a net work output. In the Otto cycle, heat transfer occurs along path DA. If no heat transfer occurs, then the return path is the same, and the net work output is zero. The lower the temperature on the path AB, the less work has to be done to compress the gas. The area inside the closed path is then greater, and so the engine does more work and is thus more efficient. Similarly, the higher the temperature along path CD, the more work output there is. (See [\[link\]](#).) So efficiency is related to the temperatures of the hot and cold reservoirs. In the next section, we shall see what the absolute limit to the efficiency of a heat engine is, and how it is related to temperature.



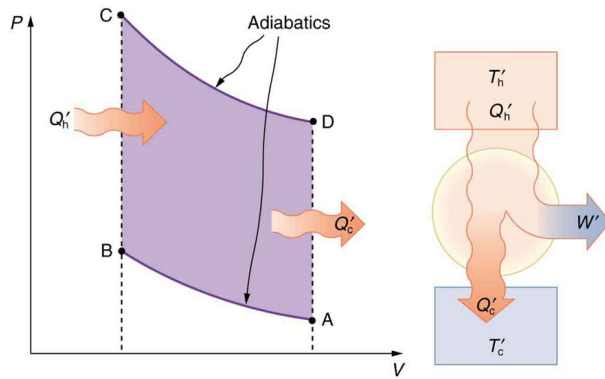
In the four-stroke internal combustion gasoline engine, heat transfer into

work takes place in the cyclical process shown here. The piston is connected to a rotating crankshaft, which both takes work out of and does work on the gas in the cylinder. (a) Air is mixed with fuel during the intake stroke. (b) During the compression stroke, the air-fuel mixture is rapidly compressed in a nearly adiabatic process, as the piston rises with the valves closed. Work is done on the gas. (c) The power stroke has two distinct parts. First, the air-fuel mixture is ignited, converting chemical potential energy into thermal energy almost instantaneously, which leads to a great increase in pressure. Then the piston descends, and the gas does work by exerting a force through a distance in a nearly adiabatic process. (d) The exhaust stroke expels the hot gas to prepare the engine for another cycle, starting again with the intake stroke.



PV diagram for a simplified Otto cycle, analogous to that employed in an internal combustion engine. Point A corresponds to the start of the compression stroke of an internal combustion engine. Paths AB and CD are adiabatic and correspond to the compression and power strokes of an internal combustion engine, respectively.

Paths BC and DA are isochoric and accomplish similar results to the ignition and exhaust-intake portions, respectively, of the internal combustion engine's cycle. Work is done on the gas along path AB, but more work is done by the gas along path CD, so that there is a net work output.



This Otto cycle produces a greater work output than the one in [\[link\]](#), because the starting temperature of path CD is higher and the starting temperature of path AB is lower. The area inside the loop is greater, corresponding to greater net work output.

Section Summary

- The two expressions of the second law of thermodynamics are: (i) Heat transfer occurs spontaneously from higher- to lower-temperature bodies but never spontaneously in the reverse direction; and (ii) It is impossible in any system for heat transfer from a reservoir to completely convert to work in a cyclical process in which the system returns to its initial state.
- Irreversible processes depend on path and do not return to their original state. Cyclical processes are processes that return to their original state at the end of every cycle.
- In a cyclical process, such as a heat engine, the net work done by the system equals the net heat transfer into the system, or $W = Q_h - Q_c$, where Q_h is the heat transfer from the hot object (hot reservoir), and Q_c is the heat transfer into the cold object (cold reservoir).

- Efficiency can be expressed as $\text{Eff} = \frac{W}{Q_h}$, the ratio of work output divided by the amount of energy input.
- The four-stroke gasoline engine is often explained in terms of the Otto cycle, which is a repeating sequence of processes that convert heat into work.

Conceptual Questions

Exercise:

Problem:

Imagine you are driving a car up Pike's Peak in Colorado. To raise a car weighing 1000 kilograms a distance of 100 meters would require about a million joules. You could raise a car 12.5 kilometers with the energy in a gallon of gas. Driving up Pike's Peak (a mere 3000-meter climb) should consume a little less than a quart of gas. But other considerations have to be taken into account. Explain, in terms of efficiency, what factors may keep you from realizing your ideal energy use on this trip.

Exercise:

Problem:

Is a temperature difference necessary to operate a heat engine? State why or why not.

Exercise:

Problem:

Definitions of efficiency vary depending on how energy is being converted. Compare the definitions of efficiency for the human body and heat engines. How does the definition of efficiency in each relate to the type of energy being converted into doing work?

Exercise:

Problem:

Why—other than the fact that the second law of thermodynamics says reversible engines are the most efficient—should heat engines employing reversible processes be more efficient than those employing irreversible processes? Consider that dissipative mechanisms are one cause of irreversibility.

Problem Exercises**Exercise:****Problem:**

A certain heat engine does 10.0 kJ of work and 8.50 kJ of heat transfer occurs to the environment in a cyclical process. (a) What was the heat transfer into this engine? (b) What was the engine's efficiency?

Solution:

(a) 18.5 kJ

(b) 54.1%

Exercise:**Problem:**

With 2.56×10^6 J of heat transfer into this engine, a given cyclical heat engine can do only 1.50×10^5 J of work. (a) What is the engine's efficiency? (b) How much heat transfer to the environment takes place?

Exercise:

Problem:

(a) What is the work output of a cyclical heat engine having a 22.0% efficiency and 6.00×10^9 J of heat transfer into the engine? (b) How much heat transfer occurs to the environment?

Solution:

(a) 1.32×10^9 J

(b) 4.68×10^9 J

Exercise:**Problem:**

(a) What is the efficiency of a cyclical heat engine in which 75.0 kJ of heat transfer occurs to the environment for every 95.0 kJ of heat transfer into the engine? (b) How much work does it produce for 100 kJ of heat transfer into the engine?

Exercise:**Problem:**

The engine of a large ship does 2.00×10^8 J of work with an efficiency of 5.00%. (a) How much heat transfer occurs to the environment? (b) How many barrels of fuel are consumed, if each barrel produces 6.00×10^9 J of heat transfer when burned?

Solution:

(a) 3.80×10^9 J

(b) 0.667 barrels

Exercise:

Problem:

(a) How much heat transfer occurs to the environment by an electrical power station that uses 1.25×10^{14} J of heat transfer into the engine with an efficiency of 42.0%? (b) What is the ratio of heat transfer to the environment to work output? (c) How much work is done?

Exercise:**Problem:**

Assume that the turbines at a coal-powered power plant were upgraded, resulting in an improvement in efficiency of 3.32%. Assume that prior to the upgrade the power station had an efficiency of 36% and that the heat transfer into the engine in one day is still the same at 2.50×10^{14} J. (a) How much more electrical energy is produced due to the upgrade? (b) How much less heat transfer occurs to the environment due to the upgrade?

Solution:

(a) 8.30×10^{12} J, which is 3.32% of 2.50×10^{14} J .

(b) -8.30×10^{12} J, where the negative sign indicates a reduction in heat transfer to the environment.

Exercise:

Problem:

This problem compares the energy output and heat transfer to the environment by two different types of nuclear power stations—one with the normal efficiency of 34.0%, and another with an improved efficiency of 40.0%. Suppose both have the same heat transfer into the engine in one day, 2.50×10^{14} J. (a) How much more electrical energy is produced by the more efficient power station? (b) How much less heat transfer occurs to the environment by the more efficient power station? (One type of more efficient nuclear power station, the gas-cooled reactor, has not been reliable enough to be economically feasible in spite of its greater efficiency.)

Glossary

irreversible process

any process that depends on path direction

second law of thermodynamics

heat transfer flows from a hotter to a cooler object, never the reverse, and some heat energy in any process is lost to available work in a cyclical process

cyclical process

a process in which the path returns to its original state at the end of every cycle

Otto cycle

a thermodynamic cycle, consisting of a pair of adiabatic processes and a pair of isochoric processes, that converts heat into work, e.g., the four-stroke engine cycle of intake, compression, ignition, and exhaust

Carnot's Perfect Heat Engine: The Second Law of Thermodynamics Restated

- Identify a Carnot cycle.
- Calculate maximum theoretical efficiency of a nuclear reactor.
- Explain how dissipative processes affect the ideal Carnot engine.



This novelty toy, known as the drinking bird, is an example of Carnot's engine. It contains methylene chloride (mixed with a dye) in the abdomen, which boils at a very low temperature—about 100°F. To operate, one gets the bird's head wet. As the water evaporates, fluid moves up into the head, causing the bird to become top-heavy and dip forward back into the water. This cools down the methylene chloride in the head, and it moves back into the abdomen, causing the bird to become bottom heavy and tip up. Except for a very small input of energy—the original head-wetting—the bird becomes a perpetual motion machine of sorts. (credit: Arabesk.nl, Wikimedia Commons)

We know from the second law of thermodynamics that a heat engine cannot be 100% efficient, since there must always be some heat transfer Q_c to the environment, which is often called waste heat. How efficient, then, can a heat engine be? This question was answered at a theoretical level in 1824 by a young French engineer, Sadi Carnot (1796–1832), in his study of the then-emerging heat engine technology crucial to the Industrial Revolution. He devised a theoretical cycle, now called the **Carnot cycle**, which is the most efficient cyclical process possible. The second law of thermodynamics can be restated in terms of the Carnot cycle, and so what Carnot actually discovered was this fundamental law. Any heat engine employing the Carnot cycle is called a **Carnot engine**.

What is crucial to the Carnot cycle—and, in fact, defines it—is that only reversible processes are used. Irreversible processes involve dissipative factors, such as friction and turbulence. This increases heat transfer Q_c to the environment and reduces the efficiency of the engine. Obviously, then, reversible processes are superior.

Note:

Carnot Engine

Stated in terms of reversible processes, the **second law of thermodynamics** has a third form:

A Carnot engine operating between two given temperatures has the greatest possible efficiency of any heat engine operating between these two temperatures. Furthermore, all engines employing only reversible processes have this same maximum efficiency when operating between the same given temperatures.

[\[link\]](#) shows the PV diagram for a Carnot cycle. The cycle comprises two isothermal and two adiabatic processes. Recall that both isothermal and adiabatic processes are, in principle, reversible.

Carnot also determined the efficiency of a perfect heat engine—that is, a Carnot engine. It is always true that the efficiency of a cyclical heat engine is given by:

Equation:

$$\text{Eff} = \frac{Q_h - Q_c}{Q_h} = 1 - \frac{Q_c}{Q_h}.$$

What Carnot found was that for a perfect heat engine, the ratio Q_c/Q_h equals the ratio of the absolute temperatures of the heat reservoirs. That is, $Q_c/Q_h = T_c/T_h$ for a Carnot engine, so that the maximum or **Carnot efficiency** Eff_C is given by

Equation:

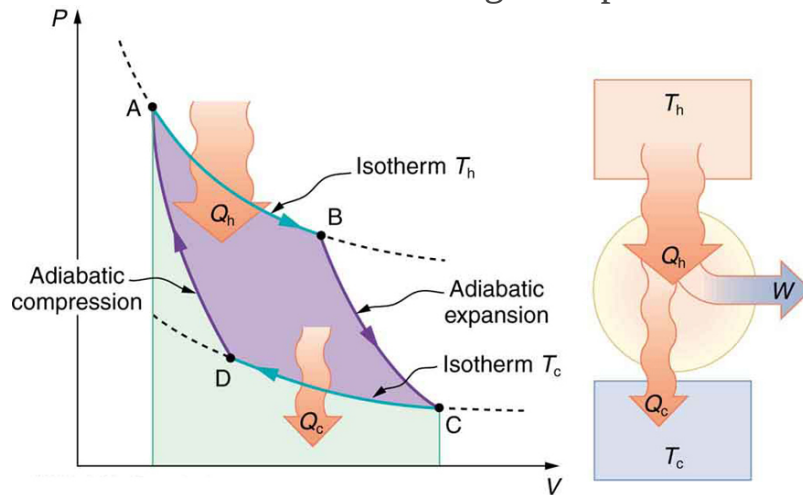
$$\text{Eff}_C = 1 - \frac{T_c}{T_h},$$

where T_h and T_c are in kelvins (or any other absolute temperature scale). No real heat engine can do as well as the Carnot efficiency—an actual efficiency of about 0.7 of this maximum is usually the best that can be accomplished. But the ideal Carnot engine, like the drinking bird above, while a fascinating novelty, has zero power. This makes it unrealistic for any applications.

Carnot's interesting result implies that 100% efficiency would be possible only if $T_c = 0$ K—that is, only if the cold reservoir were at absolute zero, a practical and theoretical impossibility. But the physical implication is this—the only way to have all heat transfer go into doing work is to remove *all* thermal energy, and this requires a cold reservoir at absolute zero.

It is also apparent that the greatest efficiencies are obtained when the ratio T_c/T_h is as small as possible. Just as discussed for the Otto cycle in the previous section, this means that efficiency is greatest for the highest possible temperature of the hot reservoir and lowest possible temperature of the cold reservoir. (This setup increases the area inside the closed loop on the PV diagram; also, it seems reasonable that the greater the temperature

difference, the easier it is to divert the heat transfer to work.) The actual reservoir temperatures of a heat engine are usually related to the type of heat source and the temperature of the environment into which heat transfer occurs. Consider the following example.

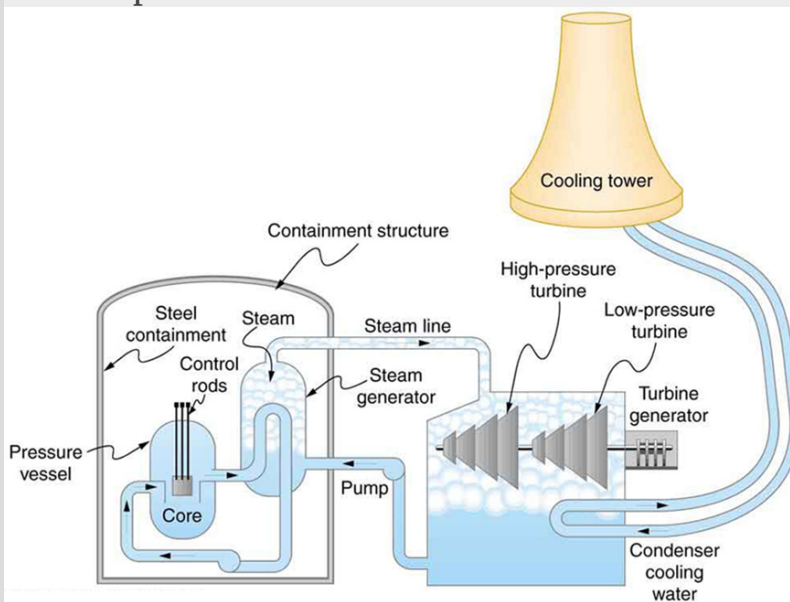


PV diagram for a Carnot cycle, employing only reversible isothermal and adiabatic processes. Heat transfer Q_h occurs into the working substance during the isothermal path AB, which takes place at constant temperature T_h . Heat transfer Q_c occurs out of the working substance during the isothermal path CD, which takes place at constant temperature T_c . The net work output W equals the area inside the path ABCDA. Also shown is a schematic of a Carnot engine operating between hot and cold reservoirs at temperatures T_h and T_c . Any heat engine using reversible processes and operating between these two temperatures will have the same maximum efficiency as the Carnot engine.

Example:

Maximum Theoretical Efficiency for a Nuclear Reactor

A nuclear power reactor has pressurized water at 300°C . (Higher temperatures are theoretically possible but practically not, due to limitations with materials used in the reactor.) Heat transfer from this water is a complex process (see [\[link\]](#)). Steam, produced in the steam generator, is used to drive the turbine-generators. Eventually the steam is condensed to water at 27°C and then heated again to start the cycle over. Calculate the maximum theoretical efficiency for a heat engine operating between these two temperatures.



Schematic diagram of a pressurized water nuclear reactor and the steam turbines that convert work into electrical energy. Heat exchange is used to generate steam, in part to avoid contamination of the generators with radioactivity. Two turbines are used because this is less expensive than operating a single generator that produces the same amount of electrical energy. The steam is condensed to liquid before being returned to the heat exchanger, to keep exit steam pressure low and aid the flow of steam through the turbines (equivalent to using a

lower-temperature cold reservoir). The considerable energy associated with condensation must be dissipated into the local environment; in this example, a cooling tower is used so there is no direct heat transfer to an aquatic environment. (Note that the water going to the cooling tower does not come into contact with the steam flowing over the turbines.)

Strategy

Since temperatures are given for the hot and cold reservoirs of this heat engine, $Eff_C = 1 - \frac{T_c}{T_h}$ can be used to calculate the Carnot (maximum theoretical) efficiency. Those temperatures must first be converted to kelvins.

Solution

The hot and cold reservoir temperatures are given as 300°C and 27.0°C , respectively. In kelvins, then, $T_h = 573\text{ K}$ and $T_c = 300\text{ K}$, so that the maximum efficiency is

Equation:

$$Eff_C = 1 - \frac{T_c}{T_h}.$$

Thus,

Equation:

$$\begin{aligned} Eff_C &= 1 - \frac{300\text{ K}}{573\text{ K}} \\ &= 0.476, \text{ or } 47.6\%. \end{aligned}$$

Discussion

A typical nuclear power station's actual efficiency is about 35%, a little better than 0.7 times the maximum possible value, a tribute to superior engineering. Electrical power stations fired by coal, oil, and natural gas have greater actual efficiencies (about 42%), because their boilers can reach higher temperatures and pressures. The cold reservoir temperature in

any of these power stations is limited by the local environment. [\[link\]](#) shows (a) the exterior of a nuclear power station and (b) the exterior of a coal-fired power station. Both have cooling towers into which water from the condenser enters the tower near the top and is sprayed downward, cooled by evaporation.



(a)



(b)

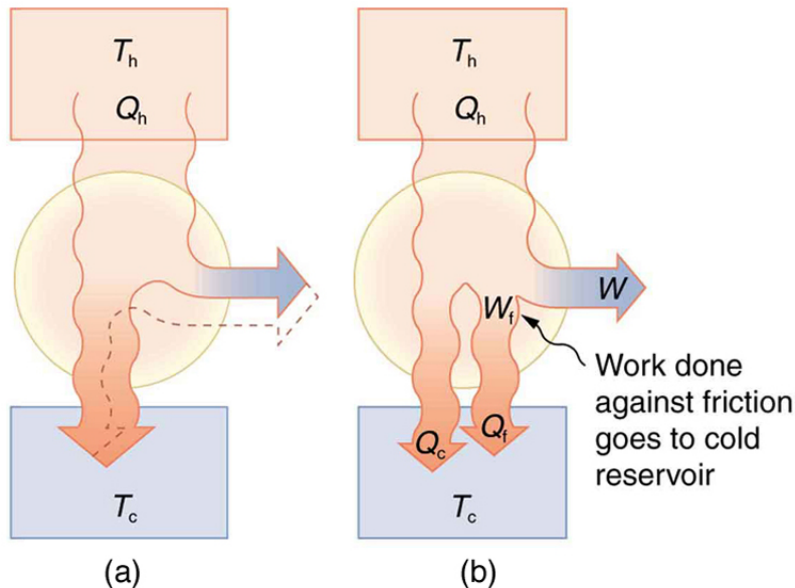
(a) A nuclear power station (credit: BlatantWorld.com) and (b) a coal-fired power station.

Both have cooling towers in which water evaporates into the environment, representing Q_c .

The nuclear reactor, which supplies Q_h , is housed inside

the dome-shaped containment buildings. (credit: Robert & Mihaela Vicol, publicphoto.org)

Since all real processes are irreversible, the actual efficiency of a heat engine can never be as great as that of a Carnot engine, as illustrated in [\[link\]](#)(a). Even with the best heat engine possible, there are always dissipative processes in peripheral equipment, such as electrical transformers or car transmissions. These further reduce the overall efficiency by converting some of the engine's work output back into heat transfer, as shown in [\[link\]](#)(b).



Real heat engines are less efficient than Carnot engines. (a) Real engines use irreversible processes, reducing the heat transfer to work. Solid lines represent the actual process; the dashed lines are what a Carnot engine would do between the same two reservoirs. (b) Friction and other dissipative processes in the output mechanisms of a heat engine convert some

of its work output into heat transfer to the environment.

Section Summary

- The Carnot cycle is a theoretical cycle that is the most efficient cyclical process possible. Any engine using the Carnot cycle, which uses only reversible processes (adiabatic and isothermal), is known as a Carnot engine.
- Any engine that uses the Carnot cycle enjoys the maximum theoretical efficiency.
- While Carnot engines are ideal engines, in reality, no engine achieves Carnot's theoretical maximum efficiency, since dissipative processes, such as friction, play a role. Carnot cycles without heat loss may be possible at absolute zero, but this has never been seen in nature.

Conceptual Questions

Exercise:

Problem:

Think about the drinking bird at the beginning of this section ([link](#)). Although the bird enjoys the theoretical maximum efficiency possible, if left to its own devices over time, the bird will cease “drinking.” What are some of the dissipative processes that might cause the bird's motion to cease?

Exercise:

Problem:

Can improved engineering and materials be employed in heat engines to reduce heat transfer into the environment? Can they eliminate heat transfer into the environment entirely?

Exercise:

Problem:

Does the second law of thermodynamics alter the conservation of energy principle?

Problem Exercises**Exercise:****Problem:**

A certain gasoline engine has an efficiency of 30.0%. What would the hot reservoir temperature be for a Carnot engine having that efficiency, if it operates with a cold reservoir temperature of 200°C?

Solution:

403°C

Exercise:**Problem:**

A gas-cooled nuclear reactor operates between hot and cold reservoir temperatures of 700°C and 27.0°C. (a) What is the maximum efficiency of a heat engine operating between these temperatures? (b) Find the ratio of this efficiency to the Carnot efficiency of a standard nuclear reactor (found in [\[link\]](#)).

Exercise:**Problem:**

(a) What is the hot reservoir temperature of a Carnot engine that has an efficiency of 42.0% and a cold reservoir temperature of 27.0°C? (b) What must the hot reservoir temperature be for a real heat engine that achieves 0.700 of the maximum efficiency, but still has an efficiency of 42.0% (and a cold reservoir at 27.0°C)? (c) Does your answer imply practical limits to the efficiency of car gasoline engines?

Solution:

(a) 244°C

(b) 477°C

(c) Yes, since automobiles engines cannot get too hot without overheating, their efficiency is limited.

Exercise:**Problem:**

Steam locomotives have an efficiency of 17.0% and operate with a hot steam temperature of 425°C. (a) What would the cold reservoir temperature be if this were a Carnot engine? (b) What would the maximum efficiency of this steam engine be if its cold reservoir temperature were 150°C?

Exercise:**Problem:**

Practical steam engines utilize 450°C steam, which is later exhausted at 270°C. (a) What is the maximum efficiency that such a heat engine can have? (b) Since 270°C steam is still quite hot, a second steam engine is sometimes operated using the exhaust of the first. What is the maximum efficiency of the second engine if its exhaust has a temperature of 150°C? (c) What is the overall efficiency of the two engines? (d) Show that this is the same efficiency as a single Carnot engine operating between 450°C and 150°C. Explicitly show how you follow the steps in the [Problem-Solving Strategies for Thermodynamics](#).

Solution:

$$(a) \text{Eff}_1 = 1 - \frac{T_{c,1}}{T_{h,1}} = 1 - \frac{543 \text{ K}}{723 \text{ K}} = 0.249 \text{ or } 24.9\%$$

$$(b) \text{Eff}_2 = 1 - \frac{423 \text{ K}}{543 \text{ K}} = 0.221 \text{ or } 22.1\%$$

$$(c) \text{Eff}_1 = 1 - \frac{T_{c,1}}{T_{h,1}} \Rightarrow T_{c,1} = T_{h,1}(1 - \text{eff}_1)$$

$$\text{similarly, } T_{c,2} = T_{h,2}(1 - \text{Eff}_2)$$

using $T_{h,2} = T_{c,1}$ in above equation gives

$$T_{c,2} = T_{h,1}(1 - \text{Eff}_1)(1 - \text{Eff}_2) \equiv T_{h,1}(1 - \text{Eff}_{\text{overall}})$$

$$\therefore (1 - \text{Eff}_{\text{overall}}) = (1 - \text{Eff}_1)(1 - \text{Eff}_2)$$

$$\text{Eff}_{\text{overall}} = 1 - (1 - 0.249)(1 - 0.221) = 41.5\%$$

$$(d) \text{Eff}_{\text{overall}} = 1 - \frac{423 \text{ K}}{723 \text{ K}} = 0.415 \text{ or } 41.5\%$$

Exercise:

Problem:

A coal-fired electrical power station has an efficiency of 38%. The temperature of the steam leaving the boiler is 550°C. What percentage of the maximum efficiency does this station obtain? (Assume the temperature of the environment is 20°C.)

Exercise:

Problem:

Would you be willing to financially back an inventor who is marketing a device that she claims has 25 kJ of heat transfer at 600 K, has heat transfer to the environment at 300 K, and does 12 kJ of work? Explain your answer.

Solution:

The heat transfer to the cold reservoir is

$$Q_c = Q_h - W = 25 \text{ kJ} - 12 \text{ kJ} = 13 \text{ kJ}, \text{ so the efficiency is}$$

$$\text{Eff} = 1 - \frac{Q_c}{Q_h} = 1 - \frac{13 \text{ kJ}}{25 \text{ kJ}} = 0.48. \text{ The Carnot efficiency is}$$

$$\text{Eff}_C = 1 - \frac{T_c}{T_h} = 1 - \frac{300 \text{ K}}{600 \text{ K}} = 0.50. \text{ The actual efficiency is } 96\%$$

of the Carnot efficiency, which is much higher than the best-ever achieved of about 70%, so her scheme is likely to be fraudulent.

Exercise:

Problem: Unreasonable Results

(a) Suppose you want to design a steam engine that has heat transfer to the environment at 270°C and has a Carnot efficiency of 0.800. What temperature of hot steam must you use? (b) What is unreasonable about the temperature? (c) Which premise is unreasonable?

Exercise:

Problem: Unreasonable Results

Calculate the cold reservoir temperature of a steam engine that uses hot steam at 450°C and has a Carnot efficiency of 0.700. (b) What is unreasonable about the temperature? (c) Which premise is unreasonable?

Solution:

(a) -56.3°C

(b) The temperature is too cold for the output of a steam engine (the local environment). It is below the freezing point of water.

(c) The assumed efficiency is too high.

Glossary

Carnot cycle

a cyclical process that uses only reversible processes, the adiabatic and isothermal processes

Carnot engine

a heat engine that uses a Carnot cycle

Carnot efficiency

the maximum theoretical efficiency for a heat engine

Applications of Thermodynamics: Heat Pumps and Refrigerators

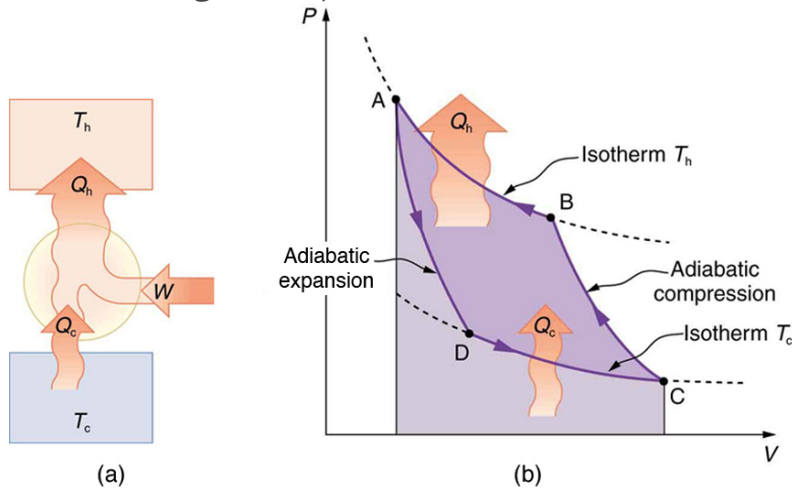
- Describe the use of heat engines in heat pumps and refrigerators.
- Demonstrate how a heat pump works to warm an interior space.
- Explain the differences between heat pumps and refrigerators.
- Calculate a heat pump's coefficient of performance.



Almost every home contains a refrigerator. Most people don't realize they are also sharing their homes with a heat pump.
(credit: Id1337x, Wikimedia Commons)

Heat pumps, air conditioners, and refrigerators utilize heat transfer from cold to hot. They are heat engines run backward. We say backward, rather than reverse, because except for Carnot engines, all heat engines, though they can be run backward, cannot truly be reversed. Heat transfer occurs from a cold reservoir Q_c and into a hot one. This requires work input W , which is also converted to heat transfer. Thus the heat transfer to the hot reservoir is $Q_h = Q_c + W$. (Note that Q_h , Q_c , and W are positive, with their directions indicated on schematics rather than by sign.) A heat pump's mission is for heat transfer Q_h to occur into a warm environment, such as a home in the winter. The mission of air conditioners and refrigerators is for

heat transfer Q_c to occur from a cool environment, such as chilling a room or keeping food at lower temperatures than the environment. (Actually, a heat pump can be used both to heat and cool a space. It is essentially an air conditioner and a heating unit all in one. In this section we will concentrate on its heating mode.)

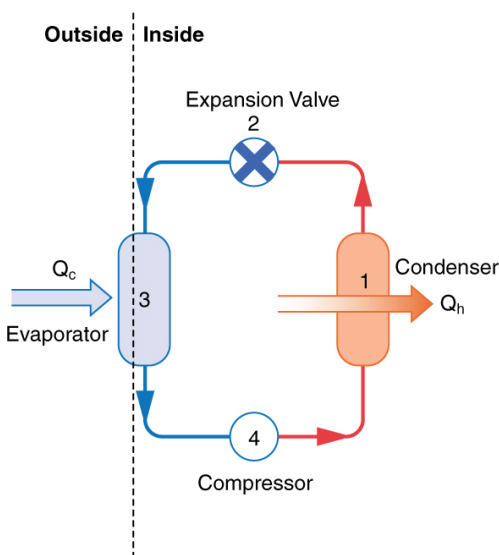


Heat pumps, air conditioners, and refrigerators are heat engines operated backward. The one shown here is based on a Carnot (reversible) engine. (a) Schematic diagram showing heat transfer from a cold reservoir to a warm reservoir with a heat pump. The directions of W , Q_h , and Q_c are opposite what they would be in a heat engine. (b) PV diagram for a Carnot cycle similar to that in [\[link\]](#) but reversed, following path ADCBA. The area inside the loop is negative, meaning there is a net work input. There is heat transfer Q_c into the system from a cold reservoir along path DC, and heat transfer Q_h out of the system into a hot reservoir along path BA.

Heat Pumps

The great advantage of using a heat pump to keep your home warm, rather than just burning fuel, is that a heat pump supplies $Q_h = Q_c + W$. Heat transfer is from the outside air, even at a temperature below freezing, to the indoor space. You only pay for W , and you get an additional heat transfer of Q_c from the outside at no cost; in many cases, at least twice as much energy is transferred to the heated space as is used to run the heat pump. When you burn fuel to keep warm, you pay for all of it. The disadvantage is that the work input (required by the second law of thermodynamics) is sometimes more expensive than simply burning fuel, especially if the work is done by electrical energy.

The basic components of a heat pump in its heating mode are shown in [\[link\]](#). A working fluid such as a non-CFC refrigerant is used. In the outdoor coils (the evaporator), heat transfer Q_c occurs to the working fluid from the cold outdoor air, turning it into a gas.



A simple heat pump has four basic components:

- (1) condenser,
- (2) expansion valve,
- (3) evaporator, and
- (4) compressor. In the

heating mode, heat transfer Q_c occurs to the working fluid in the evaporator (3) from the colder outdoor air, turning it into a gas. The electrically driven compressor (4) increases the temperature and pressure of the gas and forces it into the condenser coils (1) inside the heated space. Because the temperature of the gas is higher than the temperature in the room, heat transfer from the gas to the room occurs as the gas condenses to a liquid. The working fluid is then cooled as it flows back through an expansion valve (2) to the outdoor evaporator coils.

The electrically driven compressor (work input W) raises the temperature and pressure of the gas and forces it into the condenser coils that are inside the heated space. Because the temperature of the gas is higher than the temperature inside the room, heat transfer to the room occurs and the gas condenses to a liquid. The liquid then flows back through a pressure-reducing valve to the outdoor evaporator coils, being cooled through expansion. (In a cooling cycle, the evaporator and condenser coils exchange roles and the flow direction of the fluid is reversed.)

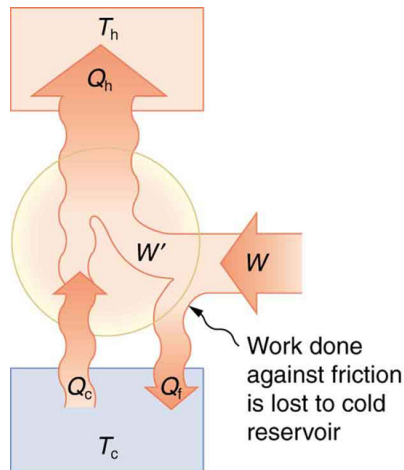
The quality of a heat pump is judged by how much heat transfer Q_h occurs into the warm space compared with how much work input W is required. In the spirit of taking the ratio of what you get to what you spend, we define a **heat pump's coefficient of performance** (COP_{hp}) to be

Equation:

$$COP_{hp} = \frac{Q_h}{W}.$$

Since the efficiency of a heat engine is $Eff = W/Q_h$, we see that $COP_{hp} = 1/Eff$, an important and interesting fact. First, since the efficiency of any heat engine is less than 1, it means that COP_{hp} is always greater than 1—that is, a heat pump always has more heat transfer Q_h than work put into it. Second, it means that heat pumps work best when temperature differences are small. The efficiency of a perfect, or Carnot, engine is $Eff_C = 1 - (T_c/T_h)$; thus, the smaller the temperature difference, the smaller the efficiency and the greater the COP_{hp} (because $COP_{hp} = 1/Eff$). In other words, heat pumps do not work as well in very cold climates as they do in more moderate climates.

Friction and other irreversible processes reduce heat engine efficiency, but they do *not* benefit the operation of a heat pump—instead, they reduce the work input by converting part of it to heat transfer back into the cold reservoir before it gets into the heat pump.



When a real heat engine is run backward, some of the intended work input (W) goes into heat transfer before it gets into the heat engine, thereby reducing its coefficient of performance COP_{hp} . In this figure, W' represents the portion of W that goes into the heat pump, while the remainder of W is lost in the form of frictional heat (Q_f) to the cold reservoir. If all of W had gone into the heat pump, then Q_h would have

been greater. The best heat pump uses adiabatic and isothermal processes, since, in theory, there would be no dissipative processes to reduce the heat transfer to the hot reservoir.

Example:

The Best COP_{hp} of a Heat Pump for Home Use

A heat pump used to warm a home must employ a cycle that produces a working fluid at temperatures greater than typical indoor temperature so that heat transfer to the inside can take place. Similarly, it must produce a working fluid at temperatures that are colder than the outdoor temperature so that heat transfer occurs from outside. Its hot and cold reservoir temperatures therefore cannot be too close, placing a limit on its COP_{hp} . (See [\[link\]](#).) What is the best coefficient of performance possible for such a heat pump, if it has a hot reservoir temperature of 45.0°C and a cold reservoir temperature of -15.0°C ?

Strategy

A Carnot engine reversed will give the best possible performance as a heat pump. As noted above, $COP_{hp} = 1/Eff$, so that we need to first calculate the Carnot efficiency to solve this problem.

Solution

Carnot efficiency in terms of absolute temperature is given by:

Equation:

$$Eff_C = 1 - \frac{T_c}{T_h}.$$

The temperatures in kelvins are $T_h = 318 \text{ K}$ and $T_c = 258 \text{ K}$, so that

Equation:

$$Eff_C = 1 - \frac{258 \text{ K}}{318 \text{ K}} = 0.1887.$$

Thus, from the discussion above,

Equation:

$$COP_{hp} = \frac{1}{Eff} = \frac{1}{0.1887} = 5.30,$$

or

Equation:

$$COP_{hp} = \frac{Q_h}{W} = 5.30,$$

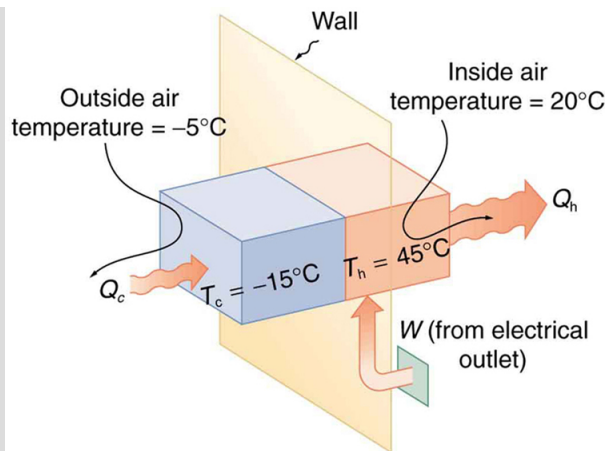
so that

Equation:

$$Q_h = 5.30 \text{ W}.$$

Discussion

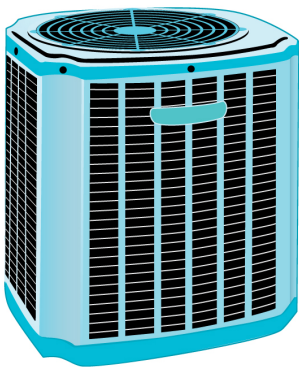
This result means that the heat transfer by the heat pump is 5.30 times as much as the work put into it. It would cost 5.30 times as much for the same heat transfer by an electric room heater as it does for that produced by this heat pump. This is not a violation of conservation of energy. Cold ambient air provides 4.3 J per 1 J of work from the electrical outlet.



Heat transfer from the outside to the inside, along with work done to run the pump, takes place in the heat pump of the example above. Note that the cold temperature produced by the heat pump is lower than the outside temperature, so that heat transfer into the working fluid occurs. The pump's compressor produces a temperature greater than the indoor temperature in order for heat transfer into the house to occur.

Real heat pumps do not perform quite as well as the ideal one in the previous example; their values of COP_{hp} range from about 2 to 4. This range means that the heat transfer Q_h from the heat pumps is 2 to 4 times as great as the work W put into them. Their economical feasibility is still limited, however, since W is usually supplied by electrical energy that costs more per joule than heat transfer by burning fuels like natural gas. Furthermore, the initial cost of a heat pump is greater than that of many

furnaces, so that a heat pump must last longer for its cost to be recovered. Heat pumps are most likely to be economically superior where winter temperatures are mild, electricity is relatively cheap, and other fuels are relatively expensive. Also, since they can cool as well as heat a space, they have advantages where cooling in summer months is also desired. Thus some of the best locations for heat pumps are in warm summer climates with cool winters. [\[link\]](#) shows a heat pump, called a “*reverse cycle*” or “*split-system cooler*” in some countries.



In hot weather, heat transfer occurs from air inside the room to air outside, cooling the room. In cool weather, heat transfer occurs from air outside to air inside, warming the room. This switching is

achieved by
reversing the
direction of
flow of the
working
fluid.

Air Conditioners and Refrigerators

Air conditioners and refrigerators are designed to cool something down in a warm environment. As with heat pumps, work input is required for heat transfer from cold to hot, and this is expensive. The quality of air conditioners and refrigerators is judged by how much heat transfer Q_c occurs from a cold environment compared with how much work input W is required. What is considered the benefit in a heat pump is considered waste heat in a refrigerator. We thus define the **coefficient of performance** (COP_{ref}) of an air conditioner or refrigerator to be

Equation:

$$COP_{\text{ref}} = \frac{Q_c}{W}.$$

Noting again that $Q_h = Q_c + W$, we can see that an air conditioner will have a lower coefficient of performance than a heat pump, because $COP_{\text{hp}} = Q_h/W$ and Q_h is greater than Q_c . In this module's Problems and Exercises, you will show that

Equation:

$$COP_{\text{ref}} = COP_{\text{hp}} - 1$$

for a heat engine used as either an air conditioner or a heat pump operating between the same two temperatures. Real air conditioners and refrigerators typically do remarkably well, having values of COP_{ref} ranging from 2 to 6.

These numbers are better than the COP_{hp} values for the heat pumps mentioned above, because the temperature differences are smaller, but they are less than those for Carnot engines operating between the same two temperatures.

A type of COP rating system called the “energy efficiency rating” (EER) has been developed. This rating is an example where non-SI units are still used and relevant to consumers. To make it easier for the consumer, Australia, Canada, New Zealand, and the U.S. use an Energy Star Rating out of 5 stars—the more stars, the more energy efficient the appliance. *EERs* are expressed in mixed units of British thermal units (Btu) per hour of heating or cooling divided by the power input in watts. Room air conditioners are readily available with *EERs* ranging from 6 to 12. Although not the same as the *COPs* just described, these *EERs* are good for comparison purposes—the greater the EER, the cheaper an air conditioner is to operate (but the higher its purchase price is likely to be).

The *EER* of an air conditioner or refrigerator can be expressed as

Equation:

$$EER = \frac{Q_c/t_1}{W/t_2},$$

where Q_c is the amount of heat transfer from a cold environment in British thermal units, t_1 is time in hours, W is the work input in joules, and t_2 is time in seconds.

Note:

Problem-Solving Strategies for Thermodynamics

1. *Examine the situation to determine whether heat, work, or internal energy are involved.* Look for any system where the primary methods of transferring energy are heat and work. Heat engines, heat pumps, refrigerators, and air conditioners are examples of such systems.

2. *Identify the system of interest and draw a labeled diagram of the system showing energy flow.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful. Maximum efficiency means a Carnot engine is involved. Efficiency is not the same as the coefficient of performance.*
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns). Be sure to distinguish heat transfer into a system from heat transfer out of the system, as well as work input from work output. In many situations, it is useful to determine the type of process, such as isothermal or adiabatic.*
5. *Solve the appropriate equation for the quantity to be determined (the unknown).*
6. *Substitute the known quantities along with their units into the appropriate equation and obtain numerical solutions complete with units.*
7. *Check the answer to see if it is reasonable: Does it make sense? For example, efficiency is always less than 1, whereas coefficients of performance are greater than 1.*

Section Summary

- An artifact of the second law of thermodynamics is the ability to heat an interior space using a heat pump. Heat pumps compress cold ambient air and, in so doing, heat it to room temperature without violation of conservation principles.
- To calculate the heat pump's coefficient of performance, use the equation $COP_{hp} = \frac{Q_h}{W}$.
- A refrigerator is a heat pump; it takes warm ambient air and expands it to chill it.

Conceptual Questions

Exercise:

Problem:

Explain why heat pumps do not work as well in very cold climates as they do in milder ones. Is the same true of refrigerators?

Exercise:**Problem:**

In some Northern European nations, homes are being built without heating systems of any type. They are very well insulated and are kept warm by the body heat of the residents. However, when the residents are not at home, it is still warm in these houses. What is a possible explanation?

Exercise:**Problem:**

Why do refrigerators, air conditioners, and heat pumps operate most cost-effectively for cycles with a small difference between T_h and T_c ? (Note that the temperatures of the cycle employed are crucial to its COP.)

Exercise:**Problem:**

Grocery store managers contend that there is *less* total energy consumption in the summer if the store is kept at a *low* temperature. Make arguments to support or refute this claim, taking into account that there are numerous refrigerators and freezers in the store.

Exercise:**Problem:**

Can you cool a kitchen by leaving the refrigerator door open?

Problem Exercises

Exercise:**Problem:**

What is the coefficient of performance of an ideal heat pump that has heat transfer from a cold temperature of -25.0°C to a hot temperature of 40.0°C ?

Solution:

4.82

Exercise:**Problem:**

Suppose you have an ideal refrigerator that cools an environment at -20.0°C and has heat transfer to another environment at 50.0°C . What is its coefficient of performance?

Exercise:**Problem:**

What is the best coefficient of performance possible for a hypothetical refrigerator that could make liquid nitrogen at -200°C and has heat transfer to the environment at 35.0°C ?

Solution:

0.311

Exercise:**Problem:**

In a very mild winter climate, a heat pump has heat transfer from an environment at 5.00°C to one at 35.0°C . What is the best possible coefficient of performance for these temperatures? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Thermodynamics](#).

Exercise:**Problem:**

(a) What is the best coefficient of performance for a heat pump that has a hot reservoir temperature of 50.0°C and a cold reservoir temperature of -20.0°C ? (b) How much heat transfer occurs into the warm environment if $3.60 \times 10^7 \text{ J}$ of work ($10.0 \text{ kW} \cdot \text{h}$) is put into it? (c) If the cost of this work input is 10.0 cents/ $\text{kW} \cdot \text{h}$, how does its cost compare with the direct heat transfer achieved by burning natural gas at a cost of 85.0 cents per therm. (A therm is a common unit of energy for natural gas and equals $1.055 \times 10^8 \text{ J}$.)

Solution:

(a) 4.61

(b) $1.66 \times 10^8 \text{ J}$ or $3.97 \times 10^4 \text{ kcal}$

(c) To transfer $1.66 \times 10^8 \text{ J}$, heat pump costs \$1.00, natural gas costs \$1.34.

Exercise:**Problem:**

(a) What is the best coefficient of performance for a refrigerator that cools an environment at -30.0°C and has heat transfer to another environment at 45.0°C ? (b) How much work in joules must be done for a heat transfer of 4186 kJ from the cold environment? (c) What is the cost of doing this if the work costs 10.0 cents per $3.60 \times 10^6 \text{ J}$ (a kilowatt-hour)? (d) How many kJ of heat transfer occurs into the warm environment? (e) Discuss what type of refrigerator might operate between these temperatures.

Exercise:

Problem:

Suppose you want to operate an ideal refrigerator with a cold temperature of -10.0°C , and you would like it to have a coefficient of performance of 7.00. What is the hot reservoir temperature for such a refrigerator?

Solution:

27.6°C

Exercise:**Problem:**

An ideal heat pump is being considered for use in heating an environment with a temperature of 22.0°C . What is the cold reservoir temperature if the pump is to have a coefficient of performance of 12.0?

Exercise:**Problem:**

A 4-ton air conditioner removes 5.06×10^7 J (48,000 British thermal units) from a cold environment in 1.00 h. (a) What energy input in joules is necessary to do this if the air conditioner has an energy efficiency rating (*EER*) of 12.0? (b) What is the cost of doing this if the work costs 10.0 cents per 3.60×10^6 J (one kilowatt-hour)? (c) Discuss whether this cost seems realistic. Note that the energy efficiency rating (*EER*) of an air conditioner or refrigerator is defined to be the number of British thermal units of heat transfer from a cold environment per hour divided by the watts of power input.

Solution:

(a) 1.44×10^7 J

(b) 40 cents

(c) This cost seems quite realistic; it says that running an air conditioner all day would cost \$9.59 (if it ran continuously).

Exercise:

Problem:

Show that the coefficients of performance of refrigerators and heat pumps are related by $COP_{\text{ref}} = COP_{\text{hp}} - 1$.

Start with the definitions of the COP s and the conservation of energy relationship between Q_h , Q_c , and W .

Glossary

heat pump

a machine that generates heat transfer from cold to hot

coefficient of performance

for a heat pump, it is the ratio of heat transfer at the output (the hot reservoir) to the work supplied; for a refrigerator or air conditioner, it is the ratio of heat transfer from the cold reservoir to the work supplied

Entropy and the Second Law of Thermodynamics: Disorder and the Unavailability of Energy

- Define entropy and calculate the increase of entropy in a system with reversible and irreversible processes.
- Explain the expected fate of the universe in entropic terms.
- Calculate the increasing disorder of a system.



The ice in this drink is slowly melting. Eventually the liquid will reach thermal equilibrium, as predicted by the second law of thermodynamics.

(credit: Jon Sullivan, PDPhoto.org)

There is yet another way of expressing the second law of thermodynamics. This version relates to a concept called **entropy**. By examining it, we shall

see that the directions associated with the second law—heat transfer from hot to cold, for example—are related to the tendency in nature for systems to become disordered and for less energy to be available for use as work. The entropy of a system can in fact be shown to be a measure of its disorder and of the unavailability of energy to do work.

Note:

Making Connections: Entropy, Energy, and Work

Recall that the simple definition of energy is the ability to do work. Entropy is a measure of how much energy is not available to do work. Although all forms of energy are interconvertible, and all can be used to do work, it is not always possible, even in principle, to convert the entire available energy into work. That unavailable energy is of interest in thermodynamics, because the field of thermodynamics arose from efforts to convert heat to work.

We can see how entropy is defined by recalling our discussion of the Carnot engine. We noted that for a Carnot cycle, and hence for any reversible processes, $Q_c/Q_h = T_c/T_h$. Rearranging terms yields

Equation:

$$\frac{Q_c}{T_c} = \frac{Q_h}{T_h}$$

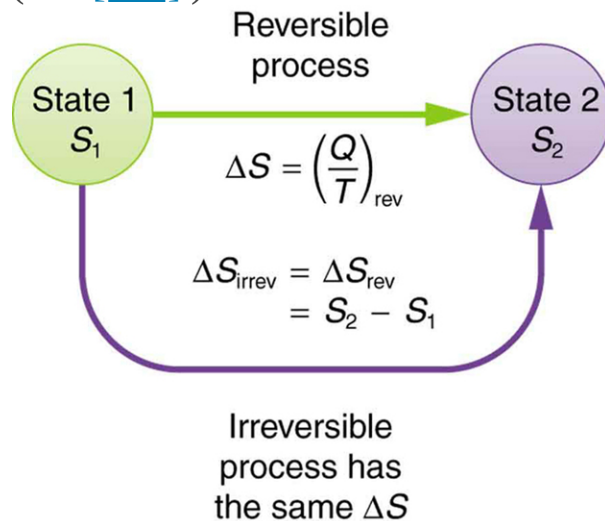
for any reversible process. Q_c and Q_h are absolute values of the heat transfer at temperatures T_c and T_h , respectively. This ratio of Q/T is defined to be the **change in entropy** ΔS for a reversible process,

Equation:

$$\Delta S = \left(\frac{Q}{T} \right)_{\text{rev}},$$

where Q is the heat transfer, which is positive for heat transfer into and negative for heat transfer out of, and T is the absolute temperature at which the reversible process takes place. The SI unit for entropy is joules per kelvin (J/K). If temperature changes during the process, then it is usually a good approximation (for small changes in temperature) to take T to be the average temperature, avoiding the need to use integral calculus to find ΔS .

The definition of ΔS is strictly valid only for reversible processes, such as used in a Carnot engine. However, we can find ΔS precisely even for real, irreversible processes. The reason is that the entropy S of a system, like internal energy U , depends only on the state of the system and not how it reached that condition. Entropy is a property of state. Thus the change in entropy ΔS of a system between state 1 and state 2 is the same no matter how the change occurs. We just need to find or imagine a reversible process that takes us from state 1 to state 2 and calculate ΔS for that process. That will be the change in entropy for any process going from state 1 to state 2. (See [\[link\]](#).)



When a system goes from state 1 to state 2, its entropy changes by the same amount ΔS , whether a hypothetical reversible path is followed or a real irreversible path is taken.

Now let us take a look at the change in entropy of a Carnot engine and its heat reservoirs for one full cycle. The hot reservoir has a loss of entropy $\Delta S_h = -Q_h/T_h$, because heat transfer occurs out of it (remember that when heat transfers out, then Q has a negative sign). The cold reservoir has a gain of entropy $\Delta S_c = Q_c/T_c$, because heat transfer occurs into it. (We assume the reservoirs are sufficiently large that their temperatures are constant.) So the total change in entropy is

Equation:

$$\Delta S_{\text{tot}} = \Delta S_h + \Delta S_c.$$

Thus, since we know that $Q_h/T_h = Q_c/T_c$ for a Carnot engine,

Equation:

$$\Delta S_{\text{tot}} = -\frac{Q_h}{T_h} + \frac{Q_c}{T_c} = 0.$$

This result, which has general validity, means that *the total change in entropy for a system in any reversible process is zero.*

The entropy of various parts of the system may change, but the total change is zero. Furthermore, the system does not affect the entropy of its surroundings, since heat transfer between them does not occur. Thus the reversible process changes neither the total entropy of the system nor the entropy of its surroundings. Sometimes this is stated as follows: *Reversible processes do not affect the total entropy of the universe.* Real processes are not reversible, though, and they do change total entropy. We can, however, use hypothetical reversible processes to determine the value of entropy in real, irreversible processes. The following example illustrates this point.

Example:

Entropy Increases in an Irreversible (Real) Process

Spontaneous heat transfer from hot to cold is an irreversible process. Calculate the total change in entropy if 4000 J of heat transfer occurs from

a hot reservoir at $T_h = 600 \text{ K}$ (327° C) to a cold reservoir at $T_c = 250 \text{ K}$ (-23° C), assuming there is no temperature change in either reservoir. (See [\[link\]](#).)

Strategy

How can we calculate the change in entropy for an irreversible process when $\Delta S_{\text{tot}} = \Delta S_h + \Delta S_c$ is valid only for reversible processes? Remember that the total change in entropy of the hot and cold reservoirs will be the same whether a reversible or irreversible process is involved in heat transfer from hot to cold. So we can calculate the change in entropy of the hot reservoir for a hypothetical reversible process in which 4000 J of heat transfer occurs from it; then we do the same for a hypothetical reversible process in which 4000 J of heat transfer occurs to the cold reservoir. This produces the same changes in the hot and cold reservoirs that would occur if the heat transfer were allowed to occur irreversibly between them, and so it also produces the same changes in entropy.

Solution

We now calculate the two changes in entropy using $\Delta S_{\text{tot}} = \Delta S_h + \Delta S_c$. First, for the heat transfer from the hot reservoir,

Equation:

$$\Delta S_h = \frac{-Q_h}{T_h} = \frac{-4000 \text{ J}}{600 \text{ K}} = -6.67 \text{ J/K}.$$

And for the cold reservoir,

Equation:

$$\Delta S_c = \frac{Q_c}{T_c} = \frac{4000 \text{ J}}{250 \text{ K}} = 16.0 \text{ J/K}.$$

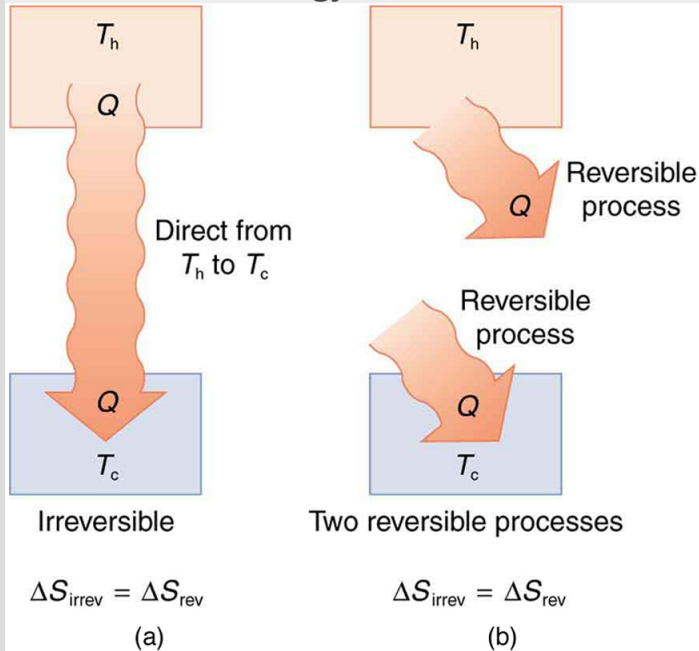
Thus the total is

Equation:

$$\begin{aligned}\Delta S_{\text{tot}} &= \Delta S_h + \Delta S_c \\ &= (-6.67 + 16.0) \text{ J/K} \\ &= 9.33 \text{ J/K}.\end{aligned}$$

Discussion

There is an *increase* in entropy for the system of two heat reservoirs undergoing this irreversible heat transfer. We will see that this means there is a loss of ability to do work with this transferred energy. Entropy has increased, and energy has become unavailable to do work.



(a) Heat transfer from a hot object to a cold one is an irreversible process that produces an overall increase in entropy. (b) The same final state and, thus, the same change in entropy is achieved for the objects if reversible heat transfer processes occur between the two objects whose temperatures are the same as the temperatures of the corresponding objects in the irreversible process.

It is reasonable that entropy increases for heat transfer from hot to cold. Since the change in entropy is Q/T , there is a larger change at lower

temperatures. The decrease in entropy of the hot object is therefore less than the increase in entropy of the cold object, producing an overall increase, just as in the previous example. This result is very general:

There is an increase in entropy for any system undergoing an irreversible process.

With respect to entropy, there are only two possibilities: entropy is constant for a reversible process, and it increases for an irreversible process. There is a fourth version of **the second law of thermodynamics stated in terms of entropy**:

The total entropy of a system either increases or remains constant in any process; it never decreases.

For example, heat transfer cannot occur spontaneously from cold to hot, because entropy would decrease.

Entropy is very different from energy. Entropy is *not* conserved but increases in all real processes. Reversible processes (such as in Carnot engines) are the processes in which the most heat transfer to work takes place and are also the ones that keep entropy constant. Thus we are led to make a connection between entropy and the availability of energy to do work.

Entropy and the Unavailability of Energy to Do Work

What does a change in entropy mean, and why should we be interested in it? One reason is that entropy is directly related to the fact that not all heat transfer can be converted into work. The next example gives some indication of how an increase in entropy results in less heat transfer into work.

Example:

Less Work is Produced by a Given Heat Transfer When Entropy Change is Greater

(a) Calculate the work output of a Carnot engine operating between temperatures of 600 K and 100 K for 4000 J of heat transfer to the engine.
(b) Now suppose that the 4000 J of heat transfer occurs first from the 600 K reservoir to a 250 K reservoir (without doing any work, and this produces the increase in entropy calculated above) before transferring into a Carnot engine operating between 250 K and 100 K. What work output is produced? (See [\[link\]](#).)

Strategy

In both parts, we must first calculate the Carnot efficiency and then the work output.

Solution (a)

The Carnot efficiency is given by

Equation:

$$Eff_C = 1 - \frac{T_c}{T_h}.$$

Substituting the given temperatures yields

Equation:

$$Eff_C = 1 - \frac{100 \text{ K}}{600 \text{ K}} = 0.833.$$

Now the work output can be calculated using the definition of efficiency for any heat engine as given by

Equation:

$$Eff = \frac{W}{Q_h}.$$

Solving for W and substituting known terms gives

Equation:

$$\begin{aligned} W &= Eff_C Q_h \\ &= (0.833)(4000 \text{ J}) = 3333 \text{ J}. \end{aligned}$$

Solution (b)

Similarly,

Equation:

$$Eff'_C = 1 - \frac{T_c}{T'_c} = 1 - \frac{100 \text{ K}}{250 \text{ K}} = 0.600,$$

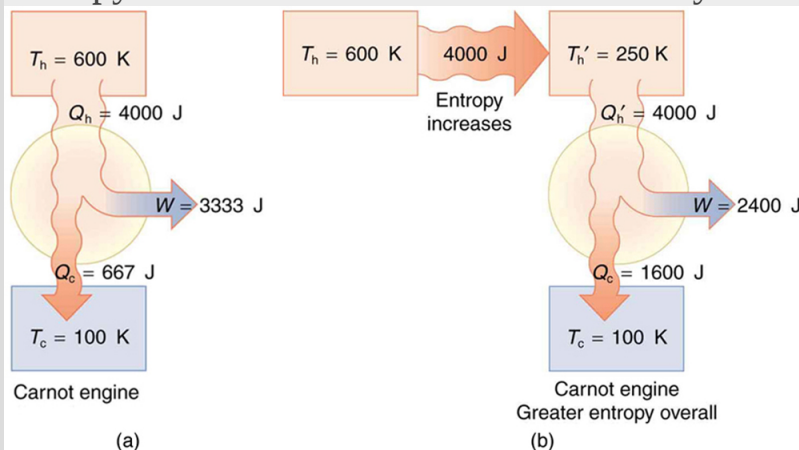
so that

Equation:

$$\begin{aligned} W &= Eff'_C Q_h \\ &= (0.600)(4000 \text{ J}) = 2400 \text{ J}. \end{aligned}$$

Discussion

There is 933 J less work from the same heat transfer in the second process. This result is important. The same heat transfer into two perfect engines produces different work outputs, because the entropy change differs in the two cases. In the second case, entropy is greater and less work is produced. Entropy is associated with the *unavailability* of energy to do work.



(a) A Carnot engine working at between 600 K and 100 K has 4000 J of heat transfer and performs 3333 J of work. (b) The 4000 J of heat transfer occurs first irreversibly to a 250 K reservoir and then goes into a Carnot engine. The increase in entropy caused by the heat transfer to a colder reservoir results in a smaller work output of 2400 J. There is a permanent loss of 933 J of energy for the purpose of doing work.

When entropy increases, a certain amount of energy becomes *permanently* unavailable to do work. The energy is not lost, but its character is changed, so that some of it can never be converted to doing work—that is, to an organized force acting through a distance. For instance, in the previous example, 933 J less work was done after an increase in entropy of 9.33 J/K occurred in the 4000 J heat transfer from the 600 K reservoir to the 250 K reservoir. It can be shown that the amount of energy that becomes unavailable for work is

Equation:

$$W_{\text{unavail}} = \Delta S \cdot T_0,$$

where T_0 is the lowest temperature utilized. In the previous example,

Equation:

$$W_{\text{unavail}} = (9.33 \text{ J/K})(100 \text{ K}) = 933 \text{ J}$$

as found.

Heat Death of the Universe: An Overdose of Entropy

In the early, energetic universe, all matter and energy were easily interchangeable and identical in nature. Gravity played a vital role in the young universe. Although it may have *seemed* disorderly, and therefore, superficially entropic, in fact, there was enormous potential energy available to do work—all the future energy in the universe.

As the universe matured, temperature differences arose, which created more opportunity for work. Stars are hotter than planets, for example, which are warmer than icy asteroids, which are warmer still than the vacuum of the space between them.

Most of these are cooling down from their usually violent births, at which time they were provided with energy of their own—nuclear energy in the case of stars, volcanic energy on Earth and other planets, and so on. Without additional energy input, however, their days are numbered.

As entropy increases, less and less energy in the universe is available to do work. On Earth, we still have great stores of energy such as fossil and nuclear fuels; large-scale temperature differences, which can provide wind energy; geothermal energies due to differences in temperature in Earth's layers; and tidal energies owing to our abundance of liquid water. As these are used, a certain fraction of the energy they contain can never be converted into doing work. Eventually, all fuels will be exhausted, all temperatures will equalize, and it will be impossible for heat engines to function, or for work to be done.

Entropy increases in a closed system, such as the universe. But in parts of the universe, for instance, in the Solar system, it is not a locally closed system. Energy flows from the Sun to the planets, replenishing Earth's stores of energy. The Sun will continue to supply us with energy for about another five billion years. We will enjoy direct solar energy, as well as side effects of solar energy, such as wind power and biomass energy from photosynthetic plants. The energy from the Sun will keep our water at the liquid state, and the Moon's gravitational pull will continue to provide tidal energy. But Earth's geothermal energy will slowly run down and won't be replenished.

But in terms of the universe, and the very long-term, very large-scale picture, the entropy of the universe is increasing, and so the availability of energy to do work is constantly decreasing. Eventually, when all stars have died, all forms of potential energy have been utilized, and all temperatures have equalized (depending on the mass of the universe, either at a very high temperature following a universal contraction, or a very low one, just before all activity ceases) there will be no possibility of doing work.

Either way, the universe is destined for thermodynamic equilibrium—maximum entropy. This is often called the *heat death of the universe*, and will mean the end of all activity. However, whether the universe contracts and heats up, or continues to expand and cools down, the end is not near.

Calculations of black holes suggest that entropy can easily continue for at least 10^{100} years.

Order to Disorder

Entropy is related not only to the unavailability of energy to do work—it is also a measure of disorder. This notion was initially postulated by Ludwig Boltzmann in the 1800s. For example, melting a block of ice means taking a highly structured and orderly system of water molecules and converting it into a disorderly liquid in which molecules have no fixed positions. (See [\[link\]](#).) There is a large increase in entropy in the process, as seen in the following example.

Example:

Entropy Associated with Disorder

Find the increase in entropy of 1.00 kg of ice originally at 0°C that is melted to form water at 0°C .

Strategy

As before, the change in entropy can be calculated from the definition of ΔS once we find the energy Q needed to melt the ice.

Solution

The change in entropy is defined as:

Equation:

$$\Delta S = \frac{Q}{T}.$$

Here Q is the heat transfer necessary to melt 1.00 kg of ice and is given by

Equation:

$$Q = mL_f,$$

where m is the mass and L_f is the latent heat of fusion. $L_f = 334\text{ kJ/kg}$ for water, so that

Equation:

$$Q = (1.00 \text{ kg})(334 \text{ kJ/kg}) = 3.34 \times 10^5 \text{ J}.$$

Now the change in entropy is positive, since heat transfer occurs into the ice to cause the phase change; thus,

Equation:

$$\Delta S = \frac{Q}{T} = \frac{3.34 \times 10^5 \text{ J}}{T}.$$

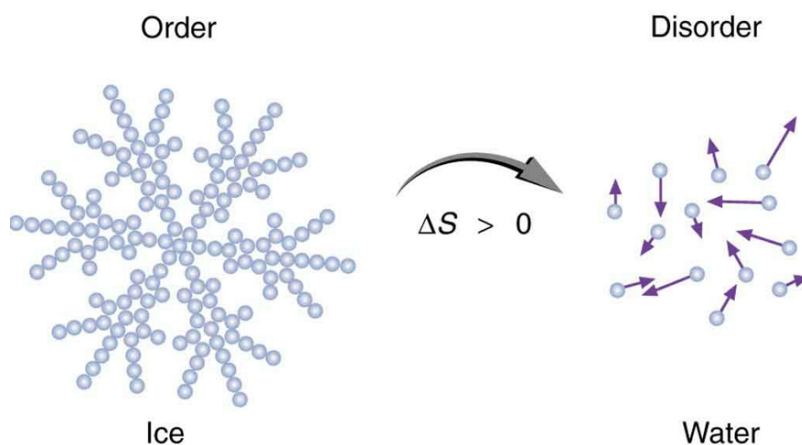
T is the melting temperature of ice. That is, $T = 0^\circ\text{C} = 273 \text{ K}$. So the change in entropy is

Equation:

$$\begin{aligned} \Delta S &= \frac{3.34 \times 10^5 \text{ J}}{273 \text{ K}} \\ &= 1.22 \times 10^3 \text{ J/K}. \end{aligned}$$

Discussion

This is a significant increase in entropy accompanying an increase in disorder.



When ice melts, it becomes more disordered and less structured. The systematic arrangement of molecules in a crystal structure is replaced by a more random and less orderly movement of molecules without

fixed locations or orientations. Its entropy increases because heat transfer occurs into it. Entropy is a measure of disorder.

In another easily imagined example, suppose we mix equal masses of water originally at two different temperatures, say 20.0°C and 40.0°C . The result is water at an intermediate temperature of 30.0°C . Three outcomes have resulted: entropy has increased, some energy has become unavailable to do work, and the system has become less orderly. Let us think about each of these results.

First, entropy has increased for the same reason that it did in the example above. Mixing the two bodies of water has the same effect as heat transfer from the hot one and the same heat transfer into the cold one. The mixing decreases the entropy of the hot water but increases the entropy of the cold water by a greater amount, producing an overall increase in entropy.

Second, once the two masses of water are mixed, there is only one temperature—you cannot run a heat engine with them. The energy that could have been used to run a heat engine is now unavailable to do work.

Third, the mixture is less orderly, or to use another term, less structured. Rather than having two masses at different temperatures and with different distributions of molecular speeds, we now have a single mass with a uniform temperature.

These three results—entropy, unavailability of energy, and disorder—are not only related but are in fact essentially equivalent.

Life, Evolution, and the Second Law of Thermodynamics

Some people misunderstand the second law of thermodynamics, stated in terms of entropy, to say that the process of the evolution of life violates this law. Over time, complex organisms evolved from much simpler ancestors, representing a large decrease in entropy of the Earth's biosphere. It is a fact

that living organisms have evolved to be highly structured, and much lower in entropy than the substances from which they grow. But it is *always* possible for the entropy of one part of the universe to decrease, provided the total change in entropy of the universe increases. In equation form, we can write this as

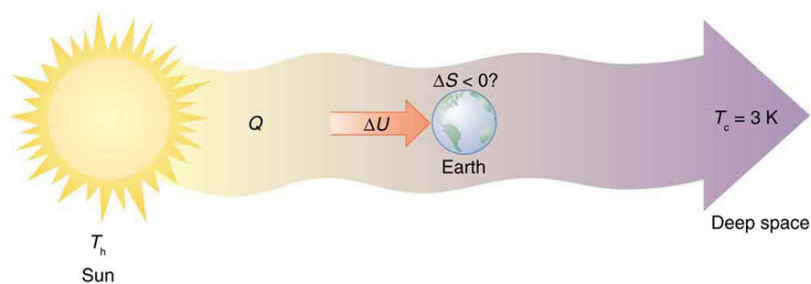
Equation:

$$\Delta S_{\text{tot}} = \Delta S_{\text{syst}} + \Delta S_{\text{envir}} > 0.$$

Thus ΔS_{syst} can be negative as long as ΔS_{envir} is positive and greater in magnitude.

How is it possible for a system to decrease its entropy? Energy transfer is necessary. If I pick up marbles that are scattered about the room and put them into a cup, my work has decreased the entropy of that system. If I gather iron ore from the ground and convert it into steel and build a bridge, my work has decreased the entropy of that system. Energy coming from the Sun can decrease the entropy of local systems on Earth—that is, ΔS_{syst} is negative. But the overall entropy of the rest of the universe increases by a greater amount—that is, ΔS_{envir} is positive and greater in magnitude. Thus, $\Delta S_{\text{tot}} = \Delta S_{\text{syst}} + \Delta S_{\text{envir}} > 0$, and the second law of thermodynamics is *not* violated.

Every time a plant stores some solar energy in the form of chemical potential energy, or an updraft of warm air lifts a soaring bird, the Earth can be viewed as a heat engine operating between a hot reservoir supplied by the Sun and a cold reservoir supplied by dark outer space—a heat engine of high complexity, causing local decreases in entropy as it uses part of the heat transfer from the Sun into deep space. There is a large total increase in entropy resulting from this massive heat transfer. A small part of this heat transfer is stored in structured systems on Earth, producing much smaller local decreases in entropy. (See [\[link\]](#).)



Earth's entropy may decrease in the process of intercepting a small part of the heat transfer from the Sun into deep space.

Entropy for the entire process increases greatly while Earth becomes more structured with living systems and stored energy in various forms.

Note:

PhET Explorations: Reversible Reactions

Watch a reaction proceed over time. How does total energy affect a reaction rate? Vary temperature, barrier height, and potential energies. Record concentrations and time in order to extract rate coefficients. Do temperature dependent studies to extract Arrhenius parameters. This simulation is best used with teacher guidance because it presents an analogy of chemical reactions. [Click to open media in new browser.](#)

Section Summary

- Entropy is the loss of energy available to do work.
- Another form of the second law of thermodynamics states that the total entropy of a system either increases or remains constant; it never decreases.

- Entropy is zero in a reversible process; it increases in an irreversible process.
- The ultimate fate of the universe is likely to be thermodynamic equilibrium, where the universal temperature is constant and no energy is available to do work.
- Entropy is also associated with the tendency toward disorder in a closed system.

Conceptual Questions

Exercise:

Problem:

A woman shuts her summer cottage up in September and returns in June. No one has entered the cottage in the meantime. Explain what she is likely to find, in terms of the second law of thermodynamics.

Exercise:

Problem:

Consider a system with a certain energy content, from which we wish to extract as much work as possible. Should the system's entropy be high or low? Is this orderly or disorderly? Structured or uniform? Explain briefly.

Exercise:

Problem:

Does a gas become more orderly when it liquefies? Does its entropy change? If so, does the entropy increase or decrease? Explain your answer.

Exercise:

Problem:

Explain how water's entropy can decrease when it freezes without violating the second law of thermodynamics. Specifically, explain what happens to the entropy of its surroundings.

Exercise:**Problem:**

Is a uniform-temperature gas more or less orderly than one with several different temperatures? Which is more structured? In which can heat transfer result in work done without heat transfer from another system?

Exercise:**Problem:**

Give an example of a spontaneous process in which a system becomes less ordered and energy becomes less available to do work. What happens to the system's entropy in this process?

Exercise:**Problem:**

What is the change in entropy in an adiabatic process? Does this imply that adiabatic processes are reversible? Can a process be precisely adiabatic for a macroscopic system?

Exercise:**Problem:**

Does the entropy of a star increase or decrease as it radiates? Does the entropy of the space into which it radiates (which has a temperature of about 3 K) increase or decrease? What does this do to the entropy of the universe?

Exercise:

Problem:

Explain why a building made of bricks has smaller entropy than the same bricks in a disorganized pile. Do this by considering the number of ways that each could be formed (the number of microstates in each macrostate).

Problem Exercises**Exercise:****Problem:**

(a) On a winter day, a certain house loses 5.00×10^8 J of heat to the outside (about 500,000 Btu). What is the total change in entropy due to this heat transfer alone, assuming an average indoor temperature of 21.0°C and an average outdoor temperature of 5.00°C ? (b) This large change in entropy implies a large amount of energy has become unavailable to do work. Where do we find more energy when such energy is lost to us?

Solution:

(a) 9.78×10^4 J/K

(b) In order to gain more energy, we must generate it from things within the house, like a heat pump, human bodies, and other appliances. As you know, we use a lot of energy to keep our houses warm in the winter because of the loss of heat to the outside.

Exercise:**Problem:**

On a hot summer day, 4.00×10^6 J of heat transfer into a parked car takes place, increasing its temperature from 35.0°C to 45.0°C . What is the increase in entropy of the car due to this heat transfer alone?

Exercise:

Problem:

A hot rock ejected from a volcano's lava fountain cools from 1100°C to 40.0°C , and its entropy decreases by 950 J/K . How much heat transfer occurs from the rock?

Solution:

$$8.01 \times 10^5 \text{ J}$$

Exercise:**Problem:**

When $1.60 \times 10^5 \text{ J}$ of heat transfer occurs into a meat pie initially at 20.0°C , its entropy increases by 480 J/K . What is its final temperature?

Exercise:**Problem:**

The Sun radiates energy at the rate of $3.80 \times 10^{26} \text{ W}$ from its 5500°C surface into dark empty space (a negligible fraction radiates onto Earth and the other planets). The effective temperature of deep space is -270°C . (a) What is the increase in entropy in one day due to this heat transfer? (b) How much work is made unavailable?

Solution:

$$(a) 1.04 \times 10^{31} \text{ J/K}$$

$$(b) 3.28 \times 10^{31} \text{ J}$$

Exercise:

Problem:

(a) In reaching equilibrium, how much heat transfer occurs from 1.00 kg of water at 40.0°C when it is placed in contact with 1.00 kg of 20.0°C water in reaching equilibrium? (b) What is the change in entropy due to this heat transfer? (c) How much work is made unavailable, taking the lowest temperature to be 20.0°C ? Explicitly show how you follow the steps in the [Problem-Solving Strategies for Entropy](#).

Exercise:**Problem:**

What is the decrease in entropy of 25.0 g of water that condenses on a bathroom mirror at a temperature of 35.0°C , assuming no change in temperature and given the latent heat of vaporization to be 2450 kJ/kg?

Solution:

199 J/K

Exercise:**Problem:**

Find the increase in entropy of 1.00 kg of liquid nitrogen that starts at its boiling temperature, boils, and warms to 20.0°C at constant pressure.

Exercise:

Problem:

A large electrical power station generates 1000 MW of electricity with an efficiency of 35.0%. (a) Calculate the heat transfer to the power station, Q_h , in one day. (b) How much heat transfer Q_c occurs to the environment in one day? (c) If the heat transfer in the cooling towers is from 35.0° C water into the local air mass, which increases in temperature from 18.0° C to 20.0° C, what is the total increase in entropy due to this heat transfer? (d) How much energy becomes unavailable to do work because of this increase in entropy, assuming an 18.0° C lowest temperature? (Part of Q_c could be utilized to operate heat engines or for simply heating the surroundings, but it rarely is.)

Solution:

(a) $2.47 \times 10^{14} \text{ J}$

(b) $1.60 \times 10^{14} \text{ J}$

(c) $2.85 \times 10^{10} \text{ J/K}$

(d) $8.29 \times 10^{12} \text{ J}$

Exercise:

Problem:

(a) How much heat transfer occurs from 20.0 kg of 90.0° C water placed in contact with 20.0 kg of 10.0° C water, producing a final temperature of 50.0° C? (b) How much work could a Carnot engine do with this heat transfer, assuming it operates between two reservoirs at constant temperatures of 90.0° C and 10.0° C? (c) What increase in entropy is produced by mixing 20.0 kg of 90.0° C water with 20.0 kg of 10.0° C water? (d) Calculate the amount of work made unavailable by this mixing using a low temperature of 10.0° C, and compare it with the work done by the Carnot engine. Explicitly show how you follow the steps in the [Problem-Solving Strategies for Entropy](#). (e) Discuss how everyday processes make increasingly more energy unavailable to do work, as implied by this problem.

Glossary**entropy**

a measurement of a system's disorder and its inability to do work in a system

change in entropy

the ratio of heat transfer to temperature Q/T

second law of thermodynamics stated in terms of entropy

the total entropy of a system either increases or remains constant; it never decreases

Statistical Interpretation of Entropy and the Second Law of Thermodynamics: The Underlying Explanation

- Identify probabilities in entropy.
- Analyze statistical probabilities in entropic systems.



When you toss a coin a large number of times, heads and tails tend to come up in roughly equal numbers. Why doesn't heads come up 100, 90, or even 80% of the time? (credit: Jon Sullivan, PDPhoto.org)

The various ways of formulating the second law of thermodynamics tell what happens rather than why it happens. Why should heat transfer occur only from hot to cold? Why should energy become ever less available to do work? Why should the universe become increasingly disorderly? The answer is that it is a matter of overwhelming probability. Disorder is simply vastly more likely than order.

When you watch an emerging rain storm begin to wet the ground, you will notice that the drops fall in a disorganized manner both in time and in space. Some fall close together, some far apart, but they never fall in

straight, orderly rows. It is not impossible for rain to fall in an orderly pattern, just highly unlikely, because there are many more disorderly ways than orderly ones. To illustrate this fact, we will examine some random processes, starting with coin tosses.

Coin Tosses

What are the possible outcomes of tossing 5 coins? Each coin can land either heads or tails. On the large scale, we are concerned only with the total heads and tails and not with the order in which heads and tails appear. The following possibilities exist:

Equation:

5 heads, 0 tails

4 heads, 1 tail

3 heads, 2 tails

2 heads, 3 tails

1 head, 4 tails

0 head, 5 tails

These are what we call macrostates. A **macrostate** is an overall property of a system. It does not specify the details of the system, such as the order in which heads and tails occur or which coins are heads or tails.

Using this nomenclature, a system of 5 coins has the 6 possible macrostates just listed. Some macrostates are more likely to occur than others. For instance, there is only one way to get 5 heads, but there are several ways to get 3 heads and 2 tails, making the latter macrostate more probable. [\[link\]](#) lists of all the ways in which 5 coins can be tossed, taking into account the order in which heads and tails occur. Each sequence is called a **microstate**—a detailed description of every element of a system.

	Individual microstates	Number of microstates
5 heads, 0 tails	HHHHH	1
4 heads, 1 tail	HHHHT, HHHTH, HHTHH, HTHHH, THHHH	5
3 heads, 2 tails	HTHTH, THTHH, HTHHT, THHTH, THHHT HTHTH, THTHH, HTHHT, THHTH, THHHT	10
2 heads, 3 tails	TTTHH, TTHHT, THHTT, HHTTT, TTHTH, THTHT, HTHTT, THTTH, HTTHT, HTTTH	10
1 head, 4 tails	TTTTH, TTTHT, TTHTT, THTTT, HTTTT	5
0 heads, 5 tails	TTTTT	1
		Total: 32

5-Coin Toss

The macrostate of 3 heads and 2 tails can be achieved in 10 ways and is thus 10 times more probable than the one having 5 heads. Not surprisingly, it is equally probable to have the reverse, 2 heads and 3 tails. Similarly, it is equally probable to get 5 tails as it is to get 5 heads. Note that all of these conclusions are based on the crucial assumption that each microstate is equally probable. With coin tosses, this requires that the coins not be

asymmetric in a way that favors one side over the other, as with loaded dice. With any system, the assumption that all microstates are equally probable must be valid, or the analysis will be erroneous.

The two most orderly possibilities are 5 heads or 5 tails. (They are more structured than the others.) They are also the least likely, only 2 out of 32 possibilities. The most disorderly possibilities are 3 heads and 2 tails and its reverse. (They are the least structured.) The most disorderly possibilities are also the most likely, with 20 out of 32 possibilities for the 3 heads and 2 tails and its reverse. If we start with an orderly array like 5 heads and toss the coins, it is very likely that we will get a less orderly array as a result, since 30 out of the 32 possibilities are less orderly. So even if you start with an orderly state, there is a strong tendency to go from order to disorder, from low entropy to high entropy. The reverse can happen, but it is unlikely.

Macrostate		Number of microstates
Heads	Tails	(W)
100	0	1
99	1	1.0×10^2
95	5	7.5×10^7
90	10	1.7×10^{13}

Macrostate		Number of microstates
75	25	2.4×10^{23}
60	40	1.4×10^{28}
55	45	6.1×10^{28}
51	49	9.9×10^{28}
50	50	1.0×10^{29}
49	51	9.9×10^{28}
45	55	6.1×10^{28}
40	60	1.4×10^{28}
25	75	2.4×10^{23}

Macrostate		Number of microstates
10	90	1.7×10^{13}
5	95	7.5×10^7
1	99	1.0×10^2
0	100	1
		Total: 1.27×10^{30}

100-Coin Toss

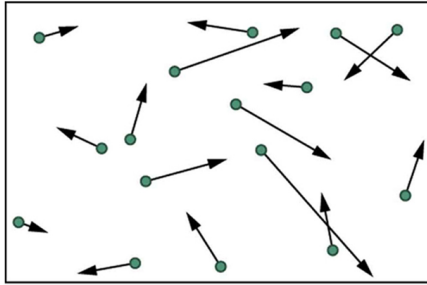
This result becomes dramatic for larger systems. Consider what happens if you have 100 coins instead of just 5. The most orderly arrangements (most structured) are 100 heads or 100 tails. The least orderly (least structured) is that of 50 heads and 50 tails. There is only 1 way (1 microstate) to get the most orderly arrangement of 100 heads. There are 100 ways (100 microstates) to get the next most orderly arrangement of 99 heads and 1 tail (also 100 to get its reverse). And there are 1.0×10^{29} ways to get 50 heads and 50 tails, the least orderly arrangement. [\[link\]](#) is an abbreviated list of the various macrostates and the number of microstates for each macrostate. The total number of microstates—the total number of different ways 100 coins can be tossed—is an impressively large 1.27×10^{30} . Now, if we start with an orderly macrostate like 100 heads and toss the coins, there is a virtual certainty that we will get a less orderly macrostate. If we keep tossing the coins, it is possible, but exceedingly unlikely, that we will ever

get back to the most orderly macrostate. If you tossed the coins once each second, you could expect to get either 100 heads or 100 tails once in 2×10^{22} years! This period is 1 trillion (10^{12}) times longer than the age of the universe, and so the chances are essentially zero. In contrast, there is an 8% chance of getting 50 heads, a 73% chance of getting from 45 to 55 heads, and a 96% chance of getting from 40 to 60 heads. Disorder is highly likely.

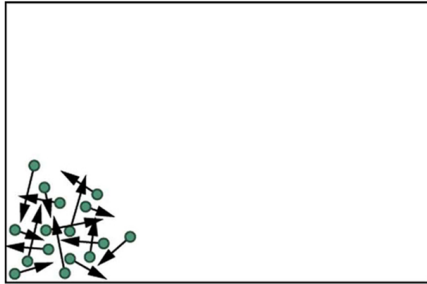
Disorder in a Gas

The fantastic growth in the odds favoring disorder that we see in going from 5 to 100 coins continues as the number of entities in the system increases. Let us now imagine applying this approach to perhaps a small sample of gas. Because counting microstates and macrostates involves statistics, this is called **statistical analysis**. The macrostates of a gas correspond to its macroscopic properties, such as volume, temperature, and pressure; and its microstates correspond to the detailed description of the positions and velocities of its atoms. Even a small amount of gas has a huge number of atoms: 1.0 cm^3 of an ideal gas at 1.0 atm and 0° C has 2.7×10^{19} atoms. So each macrostate has an immense number of microstates. In plain language, this means that there are an immense number of ways in which the atoms in a gas can be arranged, while still having the same pressure, temperature, and so on.

The most likely conditions (or macrostates) for a gas are those we see all the time—a random distribution of atoms in space with a Maxwell-Boltzmann distribution of speeds in random directions, as predicted by kinetic theory. This is the most disorderly and least structured condition we can imagine. In contrast, one type of very orderly and structured macrostate has all of the atoms in one corner of a container with identical velocities. There are very few ways to accomplish this (very few microstates corresponding to it), and so it is exceedingly unlikely ever to occur. (See [\[link\]](#)(b).) Indeed, it is so unlikely that we have a law saying that it is impossible, which has never been observed to be violated—the second law of thermodynamics.



(a) Likely



(b) Highly unlikely

(a) The ordinary state of gas in a container is a disorderly, random distribution of atoms or molecules with a Maxwell-Boltzmann distribution of speeds. It is so unlikely that these atoms or molecules would ever end up in one corner of the container that it might as well be impossible. (b) With energy transfer, the gas can be forced into one corner and its entropy greatly reduced. But left alone, it will

spontaneously
increase its entropy
and return to the
normal conditions,
because they are
immensely more
likely.

The disordered condition is one of high entropy, and the ordered one has low entropy. With a transfer of energy from another system, we could force all of the atoms into one corner and have a local decrease in entropy, but at the cost of an overall increase in entropy of the universe. If the atoms start out in one corner, they will quickly disperse and become uniformly distributed and will never return to the orderly original state ([\[link\]](#)(b)). Entropy will increase. With such a large sample of atoms, it is possible—but unimaginably unlikely—for entropy to decrease. Disorder is vastly more likely than order.

The arguments that disorder and high entropy are the most probable states are quite convincing. The great Austrian physicist Ludwig Boltzmann (1844–1906)—who, along with Maxwell, made so many contributions to kinetic theory—proved that the entropy of a system in a given state (a macrostate) can be written as

Equation:

$$S = k \ln W,$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate. W is proportional to the probability that the macrostate will occur. Thus entropy is directly related to the probability of a state—the more likely the state, the greater its entropy. Boltzmann proved that this expression for S is equivalent to the definition $\Delta S = Q/T$, which we have used extensively.

Thus the second law of thermodynamics is explained on a very basic level: entropy either remains the same or increases in every process. This phenomenon is due to the extraordinarily small probability of a decrease, based on the extraordinarily larger number of microstates in systems with greater entropy. Entropy *can* decrease, but for any macroscopic system, this outcome is so unlikely that it will never be observed.

Example:**Entropy Increases in a Coin Toss**

Suppose you toss 100 coins starting with 60 heads and 40 tails, and you get the most likely result, 50 heads and 50 tails. What is the change in entropy?

Strategy

Noting that the number of microstates is labeled W in [\[link\]](#) for the 100-coin toss, we can use $\Delta S = S_f - S_i = k \ln W_f - k \ln W_i$ to calculate the change in entropy.

Solution

The change in entropy is

Equation:

$$\Delta S = S_f - S_i = k \ln W_f - k \ln W_i,$$

where the subscript i stands for the initial 60 heads and 40 tails state, and the subscript f for the final 50 heads and 50 tails state. Substituting the values for W from [\[link\]](#) gives

Equation:

$$\begin{aligned}\Delta S &= (1.38 \times 10^{-23} \text{ J/K}) [\ln(1.0 \times 10^{29}) - \ln(1.4 \times 10^{28})] \\ &= 2.7 \times 10^{-23} \text{ J/K}\end{aligned}$$

Discussion

This increase in entropy means we have moved to a less orderly situation. It is not impossible for further tosses to produce the initial state of 60 heads and 40 tails, but it is less likely. There is about a 1 in 90 chance for that decrease in entropy ($-2.7 \times 10^{-23} \text{ J/K}$) to occur. If we calculate the decrease in entropy to move to the most orderly state, we get

$\Delta S = -92 \times 10^{-23} \text{ J/K}$. There is about a 1 in 10^{30} chance of this change occurring. So while very small decreases in entropy are unlikely, slightly greater decreases are impossibly unlikely. These probabilities imply, again, that for a macroscopic system, a decrease in entropy is impossible. For example, for heat transfer to occur spontaneously from 1.00 kg of 0°C ice to its 0°C environment, there would be a decrease in entropy of $1.22 \times 10^3 \text{ J/K}$. Given that a ΔS of 10^{-21} J/K corresponds to about a 1 in 10^{30} chance, a decrease of this size (10^3 J/K) is an *utter* impossibility. Even for a milligram of melted ice to spontaneously refreeze is impossible.

Note:**Problem-Solving Strategies for Entropy**

1. *Examine the situation to determine if entropy is involved.*
2. *Identify the system of interest and draw a labeled diagram of the system showing energy flow.*
3. *Identify exactly what needs to be determined in the problem (identify the unknowns). A written list is useful.*
4. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns). You must carefully identify the heat transfer, if any, and the temperature at which the process takes place. It is also important to identify the initial and final states.*
5. *Solve the appropriate equation for the quantity to be determined (the unknown). Note that the change in entropy can be determined between any states by calculating it for a reversible process.*
6. *Substitute the known value along with their units into the appropriate equation, and obtain numerical solutions complete with units.*
7. *To see if it is reasonable: Does it make sense?* For example, total entropy should increase for any real process or be constant for a reversible process. Disordered states should be more probable and have greater entropy than ordered states.

Section Summary

- Disorder is far more likely than order, which can be seen statistically.
- The entropy of a system in a given state (a macrostate) can be written

as

Equation:

$$S = k \ln W,$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, and $\ln W$ is the natural logarithm of the number of microstates W corresponding to the given macrostate.

Conceptual Questions

Exercise:

Problem:

Explain why a building made of bricks has smaller entropy than the same bricks in a disorganized pile. Do this by considering the number of ways that each could be formed (the number of microstates in each macrostate).

Problem Exercises

Exercise:

Problem:

Using [\[link\]](#), verify the contention that if you toss 100 coins each second, you can expect to get 100 heads or 100 tails once in 2×10^{22} years; calculate the time to two-digit accuracy.

Solution:

It should happen twice in every 1.27×10^{30} s or once in every

$$6.35 \times 10^{29} \text{ s} \left(6.35 \times 10^{29} \text{ s} \right) \left(\frac{1 \text{ h}}{3600 \text{ s}} \right) \left(\frac{1 \text{ d}}{24 \text{ h}} \right) \left(\frac{1 \text{ y}}{365.25 \text{ d}} \right)$$

$$= 2.0 \times 10^{22} \text{ y}$$

Exercise:

Problem:

What percent of the time will you get something in the range from 60 heads and 40 tails through 40 heads and 60 tails when tossing 100 coins? The total number of microstates in that range is 1.22×10^{30} . (Consult [\[link\]](#).)

Exercise:

Problem:

(a) If tossing 100 coins, how many ways (microstates) are there to get the three most likely macrostates of 49 heads and 51 tails, 50 heads and 50 tails, and 51 heads and 49 tails? (b) What percent of the total possibilities is this? (Consult [\[link\]](#).)

Solution:

(a) 3.0×10^{29}

(b) 24%

Exercise:

Problem:

(a) What is the change in entropy if you start with 100 coins in the 45 heads and 55 tails macrostate, toss them, and get 51 heads and 49 tails? (b) What if you get 75 heads and 25 tails? (c) How much more likely is 51 heads and 49 tails than 75 heads and 25 tails? (d) Does either outcome violate the second law of thermodynamics?

Exercise:

Problem:

(a) What is the change in entropy if you start with 10 coins in the 5 heads and 5 tails macrostate, toss them, and get 2 heads and 8 tails? (b) How much more likely is 5 heads and 5 tails than 2 heads and 8 tails? (Take the ratio of the number of microstates to find out.) (c) If you were betting on 2 heads and 8 tails would you accept odds of 252 to 45? Explain why or why not.

Solution:

(a) $-2.38 \times 10^{-23} \text{ J/K}$

(b) 5.6 times more likely

(c) If you were betting on two heads and 8 tails, the odds of breaking even are 252 to 45, so on average you would break even. So, no, you wouldn't bet on odds of 252 to 45.

Macrostate		Number of Microstates (W)
Heads	Tails	
10	0	1
9	1	10
8	2	45
7	3	120

Macrostate		Number of Microstates (W)
6	4	210
5	5	252
4	6	210
3	7	120
2	8	45
1	9	10
0	10	1
		Total: 1024

10-Coin Toss

Exercise:

Problem:

(a) If you toss 10 coins, what percent of the time will you get the three most likely macrostates (6 heads and 4 tails, 5 heads and 5 tails, 4 heads and 6 tails)? (b) You can realistically toss 10 coins and count the number of heads and tails about twice a minute. At that rate, how long will it take on average to get either 10 heads and 0 tails or 0 heads and 10 tails?

Exercise:

Problem:

(a) Construct a table showing the macrostates and all of the individual microstates for tossing 6 coins. (Use [\[link\]](#) as a guide.) (b) How many macrostates are there? (c) What is the total number of microstates? (d) What percent chance is there of tossing 5 heads and 1 tail? (e) How much more likely are you to toss 3 heads and 3 tails than 5 heads and 1 tail? (Take the ratio of the number of microstates to find out.)

Solution:

(b) 7

(c) 64

(d) 9.38%

(e) 3.33 times more likely (20 to 6)

Exercise:**Problem:**

In an air conditioner, 12.65 MJ of heat transfer occurs from a cold environment in 1.00 h. (a) What mass of ice melting would involve the same heat transfer? (b) How many hours of operation would be equivalent to melting 900 kg of ice? (c) If ice costs 20 cents per kg, do you think the air conditioner could be operated more cheaply than by simply using ice? Describe in detail how you evaluate the relative costs.

Glossary

macrostate

an overall property of a system

microstate

each sequence within a larger macrostate

statistical analysis

using statistics to examine data, such as counting microstates and macrostates

Energy and the Simple Harmonic Oscillator

- Determine the maximum speed of an oscillating system.

To study the energy of a simple harmonic oscillator, we first consider all the forms of energy it can have. We know from [Hooke's Law: Stress and Strain Revisited](#) that the energy stored in the deformation of a simple harmonic oscillator is a form of potential energy given by:

Equation:

$$PE_{\text{el}} = \frac{1}{2}kx^2.$$

Because a simple harmonic oscillator has no dissipative forces, the other important form of energy is kinetic energy KE. Conservation of energy for these two forms is:

Equation:

$$KE + PE_{\text{el}} = \text{constant}$$

or

Equation:

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \text{constant}.$$

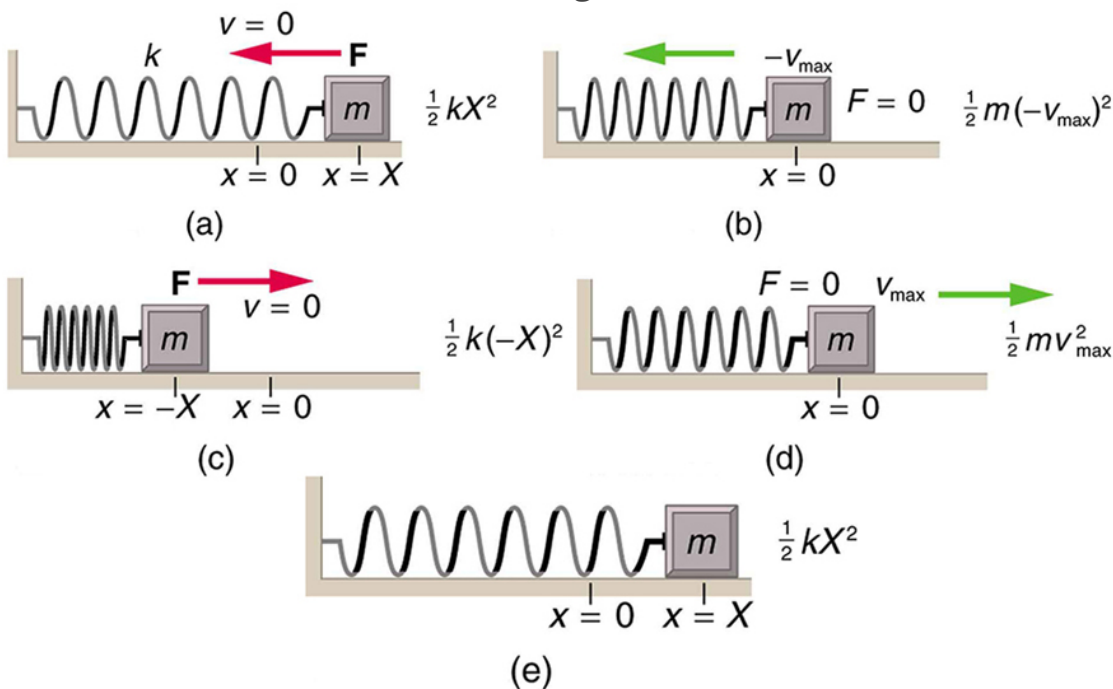
This statement of conservation of energy is valid for *all* simple harmonic oscillators, including ones where the gravitational force plays a role

Namely, for a simple pendulum we replace the velocity with $v = L\omega$, the spring constant with $k = mg/L$, and the displacement term with $x = L\theta$. Thus

Equation:

$$\frac{1}{2}mL^2\omega^2 + \frac{1}{2}mgL\theta^2 = \text{constant}.$$

In the case of undamped simple harmonic motion, the energy oscillates back and forth between kinetic and potential, going completely from one to the other as the system oscillates. So for the simple example of an object on a frictionless surface attached to a spring, as shown again in [\[link\]](#), the motion starts with all of the energy stored in the spring. As the object starts to move, the elastic potential energy is converted to kinetic energy, becoming entirely kinetic energy at the equilibrium position. It is then converted back into elastic potential energy by the spring, the velocity becomes zero when the kinetic energy is completely converted, and so on. This concept provides extra insight here and in later applications of simple harmonic motion, such as alternating current circuits.



The transformation of energy in simple harmonic motion is illustrated for an object attached to a spring on a frictionless surface.

The conservation of energy principle can be used to derive an expression for velocity v . If we start our simple harmonic motion with zero velocity and maximum displacement ($x = X$), then the total energy is

Equation:

$$\frac{1}{2}kX^2.$$

This total energy is constant and is shifted back and forth between kinetic energy and potential energy, at most times being shared by each. The conservation of energy for this system in equation form is thus:

Equation:

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{1}{2}kX^2.$$

Solving this equation for v yields:

Equation:

$$v = \pm \sqrt{\frac{k}{m}(X^2 - x^2)}.$$

Manipulating this expression algebraically gives:

Equation:

$$v = \pm \sqrt{\frac{k}{m}}X\sqrt{1 - \frac{x^2}{X^2}}$$

and so

Equation:

$$v = \pm v_{\max} \sqrt{1 - \frac{x^2}{X^2}},$$

where

Equation:

$$v_{\max} = \sqrt{\frac{k}{m}} X.$$

From this expression, we see that the velocity is a maximum ($v_{\max}$) at $x = 0$, as stated earlier in $v(t) = -v_{\max} \sin \frac{2\pi t}{T}$. Notice that the maximum velocity depends on three factors. Maximum velocity is directly proportional to amplitude. As you might guess, the greater the maximum displacement the greater the maximum velocity. Maximum velocity is also greater for stiffer systems, because they exert greater force for the same displacement. This observation is seen in the expression for $v_{\max}$; it is proportional to the square root of the force constant k . Finally, the maximum velocity is smaller for objects that have larger masses, because the maximum velocity is inversely proportional to the square root of m . For a given force, objects that have large masses accelerate more slowly.

A similar calculation for the simple pendulum produces a similar result, namely:

Equation:

$$\omega_{\max} = \sqrt{\frac{g}{L}} \theta_{\max}.$$

Example:

Determine the Maximum Speed of an Oscillating System: A Bumpy Road

Suppose that a car is 900 kg and has a suspension system that has a force constant $k = 6.53 \times 10^4$ N/m. The car hits a bump and bounces with an amplitude of 0.100 m. What is its maximum vertical velocity if you assume no damping occurs?

Strategy

We can use the expression for $v_{\max}$ given in $v_{\max} = \sqrt{\frac{k}{m}} X$ to determine the maximum vertical velocity. The variables m and k are given in the

problem statement, and the maximum displacement X is 0.100 m.

Solution

1. Identify known.

2. Substitute known values into $v_{\max} = \sqrt{\frac{k}{m}} X$:

Equation:

$$v_{\max} = \sqrt{\frac{6.53 \times 10^4 \text{ N/m}}{900 \text{ kg}}} (0.100 \text{ m}).$$

3. Calculate to find $v_{\max} = 0.852 \text{ m/s}$.

Discussion

This answer seems reasonable for a bouncing car. There are other ways to use conservation of energy to find $v_{\max}$. We could use it directly, as was done in the example featured in [Hooke's Law: Stress and Strain Revisited](#).

The small vertical displacement y of an oscillating simple pendulum, starting from its equilibrium position, is given as

Equation:

$$y(t) = a \sin \omega t,$$

where a is the amplitude, ω is the angular velocity and t is the time taken. Substituting $\omega = \frac{2\pi}{T}$, we have

Equation:

$$y(t) = a \sin \left(\frac{2\pi t}{T} \right).$$

Thus, the displacement of pendulum is a function of time as shown above. Also the velocity of the pendulum is given by

Equation:

$$v(t) = \frac{2a\pi}{T} \cos \left(\frac{2\pi t}{T} \right),$$

so the motion of the pendulum is a function of time.

Exercise:

Check Your Understanding

Problem:

Why does it hurt more if your hand is snapped with a ruler than with a loose spring, even if the displacement of each system is equal?

Solution:

The ruler is a stiffer system, which carries greater force for the same amount of displacement. The ruler snaps your hand with greater force, which hurts more.

Exercise:

Check Your Understanding

Problem:

You are observing a simple harmonic oscillator. Identify one way you could decrease the maximum velocity of the system.

Solution:

You could increase the mass of the object that is oscillating.

Section Summary

- Energy in the simple harmonic oscillator is shared between elastic potential energy and kinetic energy, with the total being constant:

Equation:

$$\frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \text{constant}.$$

- Maximum velocity depends on three factors: it is directly proportional to amplitude, it is greater for stiffer systems, and it is smaller for objects that have larger masses:

Equation:

$$v_{\max} = \sqrt{\frac{k}{m}} X.$$

Conceptual Questions

Exercise:

Problem:

Explain in terms of energy how dissipative forces such as friction reduce the amplitude of a harmonic oscillator. Also explain how a driving mechanism can compensate. (A pendulum clock is such a system.)

Problems & Exercises

Exercise:

Problem:

The length of nylon rope from which a mountain climber is suspended has a force constant of $1.40 \times 10^4 \text{ N/m}$.

(a) What is the frequency at which he bounces, given his mass plus and the mass of his equipment are 90.0 kg?

(b) How much would this rope stretch to break the climber's fall if he free-falls 2.00 m before the rope runs out of slack? Hint: Use conservation of energy. Ignore the energy the climber gains as the rope stretches.

Solution:

(a) 1.99 Hz

(b) 50.2 cm

Exercise:**Problem: Engineering Application**

Near the top of the Citigroup Center building in New York City, there is an object with mass of 4.00×10^5 kg on springs that have adjustable force constants. Its function is to dampen wind-driven oscillations of the building by oscillating at the same frequency as the building is being driven—the driving force is transferred to the object, which oscillates instead of the entire building. (a) What effective force constant should the springs have to make the object oscillate with a period of 2.00 s? (b) What energy is stored in the springs for a 2.00-m displacement from equilibrium?

Solution:

(a) 3.95×10^6 N/m

(b) 7.90×10^6 J

Waves

- State the characteristics of a wave.
- Calculate the velocity of wave propagation.



Waves in the ocean behave similarly to all other types of waves. (credit: Steve Jurveston, Flickr)

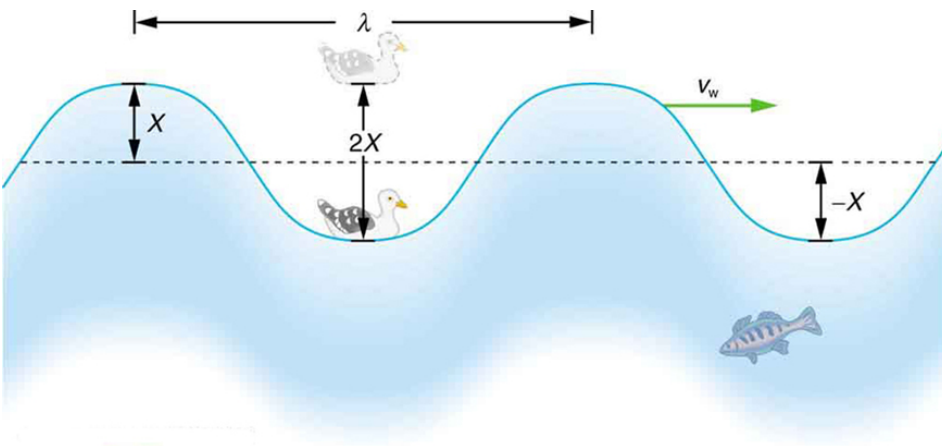
What do we mean when we say something is a wave? The most intuitive and easiest wave to imagine is the familiar water wave. More precisely, a **wave** is a disturbance that propagates, or moves from the place it was created. For water waves, the disturbance is in the surface of the water, perhaps created by a rock thrown into a pond or by a swimmer splashing the surface repeatedly. For sound waves, the disturbance is a change in air pressure, perhaps created by the oscillating cone inside a speaker. For earthquakes, there are several types of disturbances, including disturbance of Earth's surface and pressure disturbances under the surface. Even radio waves are most easily understood using an analogy with water waves. Visualizing water waves is useful because there is more to it than just a mental image. Water waves exhibit characteristics common to all waves, such as amplitude, period, frequency and energy. All wave characteristics can be described by a small set of underlying principles.

A wave is a disturbance that propagates, or moves from the place it was created. The simplest waves repeat themselves for several cycles and are associated with simple harmonic motion. Let us start by considering the simplified water wave in [\[link\]](#). The wave is an up and down disturbance of the water surface. It causes a sea gull to move up and down in simple harmonic motion as the wave crests and troughs (peaks and valleys) pass under the bird. The time for one complete up and down motion is the wave's period T . The wave's frequency is $f = 1/T$, as usual. The wave itself moves to the right in the figure. This movement of the wave is actually the disturbance moving to the right, not the water itself (or the bird would move to the right). We define **wave velocity** v_w to be the speed at which the disturbance moves. Wave velocity is sometimes also called the *propagation velocity* or *propagation speed*, because the disturbance propagates from one location to another.

Note:

Misconception Alert

Many people think that water waves push water from one direction to another. In fact, the particles of water tend to stay in one location, save for moving up and down due to the energy in the wave. The energy moves forward through the water, but the water stays in one place. If you feel yourself pushed in an ocean, what you feel is the energy of the wave, not a rush of water.



An idealized ocean wave passes under a sea gull that bobs up and down in simple harmonic motion. The wave has a wavelength λ , which is the distance between adjacent identical parts of the wave. The up and down disturbance of the surface propagates parallel to the surface at a speed v_w .

The water wave in the figure also has a length associated with it, called its **wavelength** λ , the distance between adjacent identical parts of a wave. (λ is the distance parallel to the direction of propagation.) The speed of propagation v_w is the distance the wave travels in a given time, which is one wavelength in the time of one period. In equation form, that is

Equation:

$$v_w = \frac{\lambda}{T}$$

or

Equation:

$$v_w = f\lambda.$$

This fundamental relationship holds for all types of waves. For water waves, v_w is the speed of a surface wave; for sound, v_w is the speed of sound; and for visible light, v_w is the speed of light, for example.

Note:

Take-Home Experiment: Waves in a Bowl

Fill a large bowl or basin with water and wait for the water to settle so there are no ripples. Gently drop a cork into the middle of the bowl. Estimate the wavelength and period of oscillation of the water wave that propagates away from the cork. Remove the cork from the bowl and wait

for the water to settle again. Gently drop the cork at a height that is different from the first drop. Does the wavelength depend upon how high above the water the cork is dropped?

Example:

Calculate the Velocity of Wave Propagation: Gull in the Ocean

Calculate the wave velocity of the ocean wave in [\[link\]](#) if the distance between wave crests is 10.0 m and the time for a sea gull to bob up and down is 5.00 s.

Strategy

We are asked to find v_w . The given information tells us that $\lambda = 10.0$ m and $T = 5.00$ s. Therefore, we can use $v_w = \frac{\lambda}{T}$ to find the wave velocity.

Solution

1. Enter the known values into $v_w = \frac{\lambda}{T}$:

Equation:

$$v_w = \frac{10.0 \text{ m}}{5.00 \text{ s}}.$$

2. Solve for v_w to find $v_w = 2.00$ m/s.

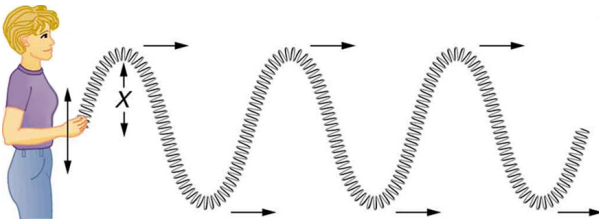
Discussion

This slow speed seems reasonable for an ocean wave. Note that the wave moves to the right in the figure at this speed, not the varying speed at which the sea gull moves up and down.

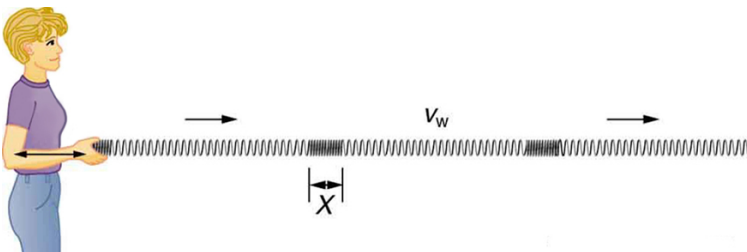
Transverse and Longitudinal Waves

A simple wave consists of a periodic disturbance that propagates from one place to another. The wave in [\[link\]](#) propagates in the horizontal direction while the surface is disturbed in the vertical direction. Such a wave is called a **transverse wave** or shear wave; in such a wave, the disturbance is perpendicular to the direction of propagation. In contrast, in a **longitudinal**

wave or compressional wave, the disturbance is parallel to the direction of propagation. [\[link\]](#) shows an example of a longitudinal wave. The size of the disturbance is its amplitude X and is completely independent of the speed of propagation v_w .



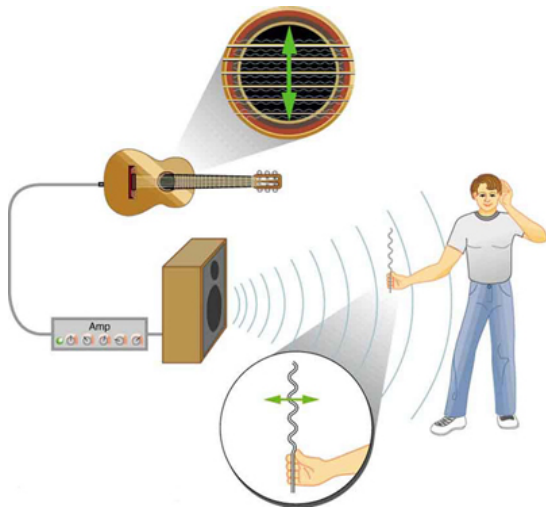
In this example of a transverse wave, the wave propagates horizontally, and the disturbance in the cord is in the vertical direction.



In this example of a longitudinal wave, the wave propagates horizontally, and the disturbance in the cord is also in the horizontal direction.

Waves may be transverse, longitudinal, or *a combination of the two*. (Water waves are actually a combination of transverse and longitudinal. The simplified water wave illustrated in [\[link\]](#) shows no longitudinal motion of the bird.) The waves on the strings of musical instruments are transverse—so are electromagnetic waves, such as visible light.

Sound waves in air and water are longitudinal. Their disturbances are periodic variations in pressure that are transmitted in fluids. Fluids do not have appreciable shear strength, and thus the sound waves in them must be longitudinal or compressional. Sound in solids can be both longitudinal and transverse.



The wave on a guitar string is transverse. The sound wave rattles a sheet of paper in a direction that shows the sound wave is longitudinal.

Earthquake waves under Earth's surface also have both longitudinal and transverse components (called compressional or P-waves and shear or S-waves, respectively). These components have important individual characteristics—they propagate at different speeds, for example.

Earthquakes also have surface waves that are similar to surface waves on water.

Exercise:

Check Your Understanding

Problem:

Why is it important to differentiate between longitudinal and transverse waves?

Solution:

In the different types of waves, energy can propagate in a different direction relative to the motion of the wave. This is important to understand how different types of waves affect the materials around them.

Note:

PhET Explorations: Wave on a String

Watch a string vibrate in slow motion. Wiggle the end of the string and make waves, or adjust the frequency and amplitude of an oscillator. Adjust the damping and tension. The end can be fixed, loose, or open.

https://phet.colorado.edu/sims/html/wave-on-a-string/latest/wave-on-a-string_en.html

Section Summary

- A wave is a disturbance that moves from the point of creation with a wave velocity v_w .
- A wave has a wavelength λ , which is the distance between adjacent identical parts of the wave.
- Wave velocity and wavelength are related to the wave's frequency and period by $v_w = \frac{\lambda}{T}$ or $v_w = f\lambda$.

- A transverse wave has a disturbance perpendicular to its direction of propagation, whereas a longitudinal wave has a disturbance parallel to its direction of propagation.

Conceptual Questions

Exercise:

Problem:

Give one example of a transverse wave and another of a longitudinal wave, being careful to note the relative directions of the disturbance and wave propagation in each.

Exercise:

Problem:

What is the difference between propagation speed and the frequency of a wave? Does one or both affect wavelength? If so, how?

Problems & Exercises

Exercise:

Problem:

Storms in the South Pacific can create waves that travel all the way to the California coast, which are 12,000 km away. How long does it take them if they travel at 15.0 m/s?

Solution:

Equation:

$$t = 9.26 \text{ d}$$

Exercise:

Problem:

Waves on a swimming pool propagate at 0.750 m/s. You splash the water at one end of the pool and observe the wave go to the opposite end, reflect, and return in 30.0 s. How far away is the other end of the pool?

Exercise:**Problem:**

Wind gusts create ripples on the ocean that have a wavelength of 5.00 cm and propagate at 2.00 m/s. What is their frequency?

Solution:**Equation:**

$$f = 40.0 \text{ Hz}$$

Exercise:**Problem:**

How many times a minute does a boat bob up and down on ocean waves that have a wavelength of 40.0 m and a propagation speed of 5.00 m/s?

Exercise:**Problem:**

Scouts at a camp shake the rope bridge they have just crossed and observe the wave crests to be 8.00 m apart. If they shake it the bridge twice per second, what is the propagation speed of the waves?

Solution:**Equation:**

$$v_w = 16.0 \text{ m/s}$$

Exercise:**Problem:**

What is the wavelength of the waves you create in a swimming pool if you splash your hand at a rate of 2.00 Hz and the waves propagate at 0.800 m/s?

Exercise:**Problem:**

What is the wavelength of an earthquake that shakes you with a frequency of 10.0 Hz and gets to another city 84.0 km away in 12.0 s?

Solution:**Equation:**

$$\lambda = 700 \text{ m}$$

Exercise:**Problem:**

Radio waves transmitted through space at $3.00 \times 10^8 \text{ m/s}$ by the Voyager spacecraft have a wavelength of 0.120 m. What is their frequency?

Exercise:**Problem:**

Your ear is capable of differentiating sounds that arrive at the ear just 1.00 ms apart. What is the minimum distance between two speakers that produce sounds that arrive at noticeably different times on a day when the speed of sound is 340 m/s?

Solution:**Equation:**

$$d = 34.0 \text{ cm}$$

Exercise:

Problem:

(a) Seismographs measure the arrival times of earthquakes with a precision of 0.100 s. To get the distance to the epicenter of the quake, they compare the arrival times of S- and P-waves, which travel at different speeds. ([link](#)) If S- and P-waves travel at 4.00 and 7.20 km/s, respectively, in the region considered, how precisely can the distance to the source of the earthquake be determined? (b) Seismic waves from underground detonations of nuclear bombs can be used to locate the test site and detect violations of test bans. Discuss whether your answer to (a) implies a serious limit to such detection. (Note also that the uncertainty is greater if there is an uncertainty in the propagation speeds of the S- and P-waves.)



A seismograph as described in above problem.(credit: Oleg Alexandrov)

Glossary

longitudinal wave

a wave in which the disturbance is parallel to the direction of propagation

transverse wave

a wave in which the disturbance is perpendicular to the direction of propagation

wave velocity

the speed at which the disturbance moves. Also called the propagation velocity or propagation speed

wavelength

the distance between adjacent identical parts of a wave

Energy in Waves: Intensity

- Calculate the intensity and the power of rays and waves.



The destructive effect of an earthquake is palpable evidence of the energy carried in these waves. The Richter scale rating of earthquakes is related to both their amplitude and the energy they carry.

(credit: Petty Officer 2nd Class Candice Villarreal, U.S. Navy)

All waves carry energy. The energy of some waves can be directly observed. Earthquakes can shake whole cities to the ground, performing the work of thousands of wrecking balls.

Loud sounds pulverize nerve cells in the inner ear, causing permanent hearing loss. Ultrasound is used for deep-heat treatment of muscle strains. A laser beam can burn away a malignancy. Water waves chew up beaches.

The amount of energy in a wave is related to its amplitude. Large-amplitude earthquakes produce large ground displacements. Loud sounds have higher pressure amplitudes and come from larger-amplitude source vibrations than

soft sounds. Large ocean breakers churn up the shore more than small ones. More quantitatively, a wave is a displacement that is resisted by a restoring force. The larger the displacement x , the larger the force $F = kx$ needed to create it. Because work W is related to force multiplied by distance (Fx) and energy is put into the wave by the work done to create it, the energy in a wave is related to amplitude. In fact, a wave's energy is directly proportional to its amplitude squared because

Equation:

$$W \propto Fx = kx^2.$$

The energy effects of a wave depend on time as well as amplitude. For example, the longer deep-heat ultrasound is applied, the more energy it transfers. Waves can also be concentrated or spread out. Sunlight, for example, can be focused to burn wood. Earthquakes spread out, so they do less damage the farther they get from the source. In both cases, changing the area the waves cover has important effects. All these pertinent factors are included in the definition of **intensity** I as power per unit area:

Equation:

$$I = \frac{P}{A}$$

where P is the power carried by the wave through area A . The definition of intensity is valid for any energy in transit, including that carried by waves. The SI unit for intensity is watts per square meter (W/m^2). For example, infrared and visible energy from the Sun impinge on Earth at an intensity of $1300 \text{ W}/\text{m}^2$ just above the atmosphere. There are other intensity-related units in use, too. The most common is the decibel. For example, a 90 decibel sound level corresponds to an intensity of $10^{-3} \text{ W}/\text{m}^2$. (This quantity is not much power per unit area considering that 90 decibels is a relatively high sound level. Decibels will be discussed in some detail in a later chapter.

Example:**Calculating intensity and power: How much energy is in a ray of sunlight?**

The average intensity of sunlight on Earth's surface is about 700 W/m^2 .

(a) Calculate the amount of energy that falls on a solar collector having an area of 0.500 m^2 in 4.00 h .

(b) What intensity would such sunlight have if concentrated by a magnifying glass onto an area 200 times smaller than its own?

Strategy a

Because power is energy per unit time or $P = \frac{E}{t}$, the definition of intensity can be written as $I = \frac{P}{A} = \frac{E/t}{A}$, and this equation can be solved for E with the given information.

Solution a

1. Begin with the equation that states the definition of intensity:

Equation:

$$I = \frac{P}{A}.$$

2. Replace P with its equivalent E/t :

Equation:

$$I = \frac{E/t}{A}.$$

3. Solve for E :

Equation:

$$E = IAt.$$

4. Substitute known values into the equation:

Equation:

$$E = (700 \text{ W/m}^2)(0.500 \text{ m}^2)[(4.00 \text{ h})(3600 \text{ s/h})].$$

5. Calculate to find E and convert units:

Equation:

$$5.04 \times 10^6 \text{ J},$$

Discussion a

The energy falling on the solar collector in 4 h in part is enough to be useful—for example, for heating a significant amount of water.

Strategy b

Taking a ratio of new intensity to old intensity and using primes for the new quantities, we will find that it depends on the ratio of the areas. All other quantities will cancel.

Solution b

1. Take the ratio of intensities, which yields:

Equation:

$$\frac{I'}{I} = \frac{P'/A'}{P/A} = \frac{A}{A'} \left(\text{The powers cancel because } P' = P \right).$$

2. Identify the knowns:

Equation:

$$A = 200A',$$

Equation:

$$\frac{I'}{I} = 200.$$

3. Substitute known quantities:

Equation:

$$I' = 200I = 200(700 \text{ W/m}^2).$$

4. Calculate to find I' :

Equation:

$$I' = 1.40 \times 10^5 \text{ W/m}^2.$$

Discussion b

Decreasing the area increases the intensity considerably. The intensity of the concentrated sunlight could even start a fire.

Example:**Determine the combined intensity of two waves: Perfect constructive interference**

If two identical waves, each having an intensity of 1.00 W/m^2 , interfere perfectly constructively, what is the intensity of the resulting wave?

Strategy

We know from [Superposition and Interference](#) that when two identical waves, which have equal amplitudes X , interfere perfectly constructively, the resulting wave has an amplitude of $2X$. Because a wave's intensity is proportional to amplitude squared, the intensity of the resulting wave is four times as great as in the individual waves.

Solution

1. Recall that intensity is proportional to amplitude squared.
2. Calculate the new amplitude:

Equation:

$$I' \propto (X')^2 = (2X)^2 = 4X^2.$$

3. Recall that the intensity of the old amplitude was:

Equation:

$$I \propto X^2.$$

4. Take the ratio of new intensity to the old intensity. This gives:

Equation:

$$\frac{I'}{I} = 4.$$

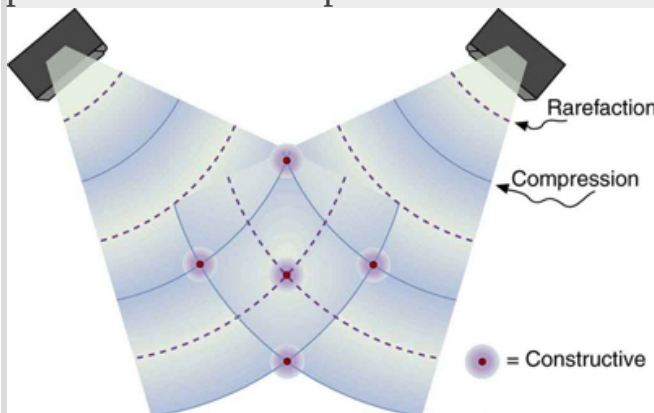
5. Calculate to find I' :

Equation:

$$I' = 4I = 4.00 \text{ W/m}^2.$$

Discussion

The intensity goes up by a factor of 4 when the amplitude doubles. This answer is a little disquieting. The two individual waves each have intensities of 1.00 W/m^2 , yet their sum has an intensity of 4.00 W/m^2 , which may appear to violate conservation of energy. This violation, of course, cannot happen. What does happen is intriguing. The area over which the intensity is 4.00 W/m^2 is much less than the area covered by the two waves before they interfered. There are other areas where the intensity is zero. The addition of waves is not as simple as our first look in [Superposition and Interference](#) suggested. We actually get a pattern of both constructive interference and destructive interference whenever two waves are added. For example, if we have two stereo speakers putting out 1.00 W/m^2 each, there will be places in the room where the intensity is 4.00 W/m^2 , other places where the intensity is zero, and others in between. [\[link\]](#) shows what this interference might look like. We will pursue interference patterns elsewhere in this text.



These stereo speakers produce both constructive interference and destructive interference in the room, a property common to the

superposition of all types of waves.
The shading is proportional to
intensity.

Exercise:

Check Your Understanding

Problem:

Which measurement of a wave is most important when determining the wave's intensity?

Solution:

Amplitude, because a wave's energy is directly proportional to its amplitude squared.

Section Summary

Intensity is defined to be the power per unit area:

$$I = \frac{P}{A} \text{ and has units of } \text{W}/\text{m}^2.$$

Conceptual Questions

Exercise:

Problem:

Two identical waves undergo pure constructive interference. Is the resultant intensity twice that of the individual waves? Explain your answer.

Exercise:

Problem:

Circular water waves decrease in amplitude as they move away from where a rock is dropped. Explain why.

Problems & Exercises**Exercise:****Problem: Medical Application**

Ultrasound of intensity $1.50 \times 10^2 \text{ W/m}^2$ is produced by the rectangular head of a medical imaging device measuring 3.00 by 5.00 cm. What is its power output?

Solution:

0.225 W

Exercise:**Problem:**

The low-frequency speaker of a stereo set has a surface area of 0.05 m^2 and produces 1W of acoustical power. What is the intensity at the speaker? If the speaker projects sound uniformly in all directions, at what distance from the speaker is the intensity 0.1 W/m^2 ?

Exercise:**Problem:**

To increase intensity of a wave by a factor of 50, by what factor should the amplitude be increased?

Solution:

7.07

Exercise:**Problem: Engineering Application**

A device called an insolation meter is used to measure the intensity of sunlight has an area of 100 cm^2 and registers 6.50 W . What is the intensity in W/m^2 ?

Exercise:**Problem: Astronomy Application**

Energy from the Sun arrives at the top of the Earth's atmosphere with an intensity of 1.30 kW/m^2 . How long does it take for $1.8 \times 10^9 \text{ J}$ to arrive on an area of 1.00 m^2 ?

Solution:

16.0 d

Exercise:**Problem:**

Suppose you have a device that extracts energy from ocean breakers in direct proportion to their intensity. If the device produces 10.0 kW of power on a day when the breakers are 1.20 m high, how much will it produce when they are 0.600 m high?

Solution:

2.50 kW

Exercise:**Problem: Engineering Application**

(a) A photovoltaic array of (solar cells) is 10.0% efficient in gathering solar energy and converting it to electricity. If the average intensity of

sunlight on one day is 700 W/m^2 , what area should your array have to gather energy at the rate of 100 W ? (b) What is the maximum cost of the array if it must pay for itself in two years of operation averaging 10.0 hours per day? Assume that it earns money at the rate of 9.00 ¢ per kilowatt-hour.

Exercise:

Problem:

A microphone receiving a pure sound tone feeds an oscilloscope, producing a wave on its screen. If the sound intensity is originally $2.00 \times 10^{-5} \text{ W/m}^2$, but is turned up until the amplitude increases by 30.0% , what is the new intensity?

Solution:

$$3.38 \times 10^{-5} \text{ W/m}^2$$

Exercise:

Problem: Medical Application

(a) What is the intensity in W/m^2 of a laser beam used to burn away cancerous tissue that, when 90.0% absorbed, puts 500 J of energy into a circular spot 2.00 mm in diameter in 4.00 s ? (b) Discuss how this intensity compares to the average intensity of sunlight (about 700 W/m^2) and the implications that would have if the laser beam entered your eye. Note how the amount of damage depends on the time duration of the exposure.

Glossary

intensity
power per unit area

Maxwell's Equations: Electromagnetic Waves Predicted and Observed

- Restate Maxwell's equations.

The Scotsman James Clerk Maxwell (1831–1879) is regarded as the greatest theoretical physicist of the 19th century. (See [\[link\]](#).) Although he died young, Maxwell not only formulated a complete electromagnetic theory, represented by **Maxwell's equations**, he also developed the kinetic theory of gases and made significant contributions to the understanding of color vision and the nature of Saturn's rings.



James Clerk Maxwell, a 19th-century physicist, developed a theory that explained the relationship between electricity and magnetism and correctly predicted that visible light is caused by electromagnetic

waves. (credit:
G. J. Stodart)

Maxwell brought together all the work that had been done by brilliant physicists such as Oersted, Coulomb, Gauss, and Faraday, and added his own insights to develop the overarching theory of electromagnetism. Maxwell's equations are paraphrased here in words because their mathematical statement is beyond the level of this text. However, the equations illustrate how apparently simple mathematical statements can elegantly unite and express a multitude of concepts—why mathematics is the language of science.

Note:

Maxwell's Equations

1. **Electric field lines** originate on positive charges and terminate on negative charges. The electric field is defined as the force per unit charge on a test charge, and the strength of the force is related to the electric constant ϵ_0 , also known as the permittivity of free space. From Maxwell's first equation we obtain a special form of Coulomb's law known as Gauss's law for electricity.
2. **Magnetic field lines** are continuous, having no beginning or end. No magnetic monopoles are known to exist. The strength of the magnetic force is related to the magnetic constant μ_0 , also known as the permeability of free space. This second of Maxwell's equations is known as Gauss's law for magnetism.
3. A changing magnetic field induces an electromotive force (emf) and, hence, an electric field. The direction of the emf opposes the change. This third of Maxwell's equations is Faraday's law of induction, and includes Lenz's law.
4. Magnetic fields are generated by moving charges or by changing electric fields. This fourth of Maxwell's equations encompasses Ampere's law and adds another source of magnetism—changing electric fields.

Maxwell's equations encompass the major laws of electricity and magnetism. What is not so apparent is the symmetry that Maxwell introduced in his mathematical framework. Especially important is his addition of the hypothesis that changing electric fields create magnetic fields. This is exactly analogous (and symmetric) to Faraday's law of induction and had been suspected for some time, but fits beautifully into Maxwell's equations.

Symmetry is apparent in nature in a wide range of situations. In contemporary research, symmetry plays a major part in the search for subatomic particles using massive multinational particle accelerators such as the new Large Hadron Collider at CERN.

Note:

Making Connections: Unification of Forces

Maxwell's complete and symmetric theory showed that electric and magnetic forces are not separate, but different manifestations of the same thing—the electromagnetic force. This classical unification of forces is one motivation for current attempts to unify the four basic forces in nature—the gravitational, electrical, strong, and weak nuclear forces.

Since changing electric fields create relatively weak magnetic fields, they could not be easily detected at the time of Maxwell's hypothesis. Maxwell realized, however, that oscillating charges, like those in AC circuits, produce changing electric fields. He predicted that these changing fields would propagate from the source like waves generated on a lake by a jumping fish.

The waves predicted by Maxwell would consist of oscillating electric and magnetic fields—defined to be an electromagnetic wave (EM wave). Electromagnetic waves would be capable of exerting forces on charges great distances from their source, and they might thus be detectable. Maxwell calculated that electromagnetic waves would propagate at a speed given by the equation

Equation:

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}.$$

When the values for μ_0 and ϵ_0 are entered into the equation for c , we find that

Equation:

$$c = \frac{1}{\sqrt{(8.85 \times 10^{-12} \frac{\text{C}^2}{\text{N} \cdot \text{m}^2})(4\pi \times 10^{-7} \frac{\text{T} \cdot \text{m}}{\text{A}})}} = 3.00 \times 10^8 \text{ m/s},$$

which is the speed of light. In fact, Maxwell concluded that light is an electromagnetic wave having such wavelengths that it can be detected by the eye.

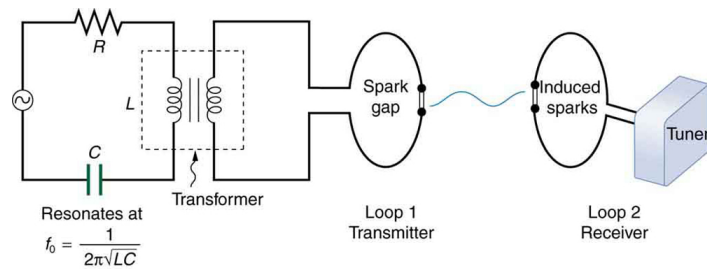
Other wavelengths should exist—it remained to be seen if they did. If so, Maxwell's theory and remarkable predictions would be verified, the greatest triumph of physics since Newton. Experimental verification came within a few years, but not before Maxwell's death.

Hertz's Observations

The German physicist Heinrich Hertz (1857–1894) was the first to generate and detect certain types of electromagnetic waves in the laboratory. Starting in 1887, he performed a series of experiments that not only confirmed the existence of electromagnetic waves, but also verified that they travel at the speed of light.

Hertz used an AC RLC (resistor-inductor-capacitor) circuit that resonates at a known frequency $f_0 = \frac{1}{2\pi\sqrt{LC}}$ and connected it to a loop of wire as shown in [\[link\]](#). High voltages induced across the gap in the loop produced sparks that were visible evidence of the current in the circuit and that helped generate electromagnetic waves.

Across the laboratory, Hertz had another loop attached to another RLC circuit, which could be tuned (as the dial on a radio) to the same resonant frequency as the first and could, thus, be made to receive electromagnetic waves. This loop also had a gap across which sparks were generated, giving solid evidence that electromagnetic waves had been received.



The apparatus used by Hertz in 1887 to generate and detect electromagnetic waves. An RLC circuit connected to the first loop caused sparks across a gap in the wire loop and generated electromagnetic waves. Sparks across a gap in the second loop located across the laboratory gave evidence that the waves had been received.

Hertz also studied the reflection, refraction, and interference patterns of the electromagnetic waves he generated, verifying their wave character. He was able to determine wavelength from the interference patterns, and knowing their frequency, he could calculate the propagation speed using the equation $v = f\lambda$ (velocity—or speed—equals frequency times wavelength). Hertz was thus able to prove that electromagnetic waves travel at the speed of light. The SI unit for frequency, the hertz ($1 \text{ Hz} = 1 \text{ cycle/s}$), is named in his honor.

Section Summary

- Electromagnetic waves consist of oscillating electric and magnetic fields and propagate at the speed of light c . They were predicted by Maxwell, who also showed that

Equation:

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}},$$

where μ_0 is the permeability of free space and ϵ_0 is the permittivity of free space.

- Maxwell's prediction of electromagnetic waves resulted from his formulation of a complete and symmetric theory of electricity and magnetism, known as Maxwell's equations.
- These four equations are paraphrased in this text, rather than presented numerically, and encompass the major laws of electricity and magnetism. First is Gauss's law for electricity, second is Gauss's law for magnetism, third is Faraday's law of induction, including Lenz's law, and fourth is Ampere's law in a symmetric formulation that adds another source of magnetism—changing electric fields.

Problems & Exercises

Exercise:

Problem:

Verify that the correct value for the speed of light c is obtained when numerical values for the permeability and permittivity of free space (μ_0 and ϵ_0) are entered into the equation $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$.

Exercise:

Problem:

Show that, when SI units for μ_0 and ϵ_0 are entered, the units given by the right-hand side of the equation in the problem above are m/s.

Glossary

electromagnetic waves

radiation in the form of waves of electric and magnetic energy

Maxwell's equations

a set of four equations that comprise a complete, overarching theory of electromagnetism

RLC circuit

an electric circuit that includes a resistor, capacitor and inductor

hertz

an SI unit denoting the frequency of an electromagnetic wave, in cycles per second

speed of light

in a vacuum, such as space, the speed of light is a constant 3×10^8 m/s

electromotive force (emf)

energy produced per unit charge, drawn from a source that produces an electrical current

electric field lines

a pattern of imaginary lines that extend between an electric source and charged objects in the surrounding area, with arrows pointed away from positively charged objects and toward negatively charged objects. The more lines in the pattern, the stronger the electric field in that region

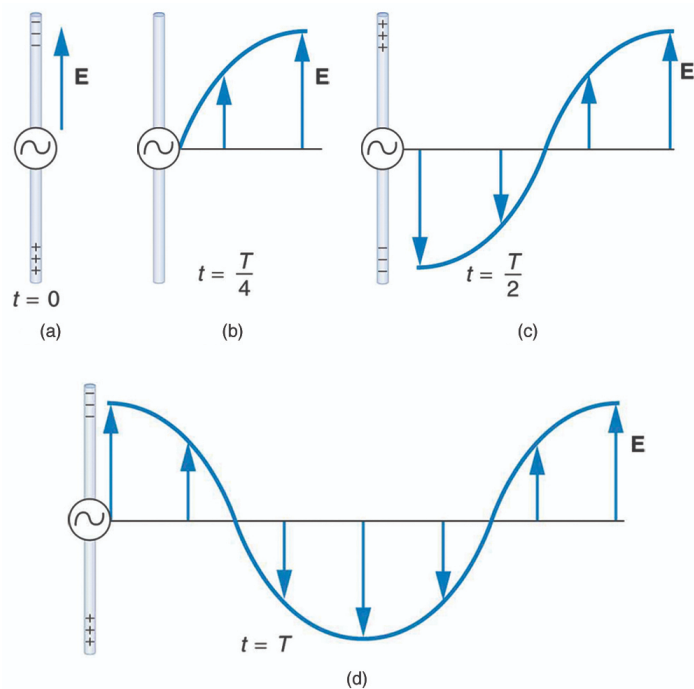
magnetic field lines

a pattern of continuous, imaginary lines that emerge from and enter into opposite magnetic poles. The density of the lines indicates the magnitude of the magnetic field

Production of Electromagnetic Waves

- Describe the electric and magnetic waves as they move out from a source, such as an AC generator.
- Explain the mathematical relationship between the magnetic field strength and the electrical field strength.
- Calculate the maximum strength of the magnetic field in an electromagnetic wave, given the maximum electric field strength.

We can get a good understanding of **electromagnetic waves** (EM) by considering how they are produced. Whenever a current varies, associated electric and magnetic fields vary, moving out from the source like waves. Perhaps the easiest situation to visualize is a varying current in a long straight wire, produced by an AC generator at its center, as illustrated in [\[link\]](#).



This long straight gray wire with an AC generator at its center becomes a broadcast antenna for electromagnetic waves. Shown here are the charge distributions at four different times. The electric field (**E**) propagates away

from the antenna at the speed of light,
forming part of an electromagnetic
wave.

The **electric field** (**E**) shown surrounding the wire is produced by the charge distribution on the wire. Both the **E** and the charge distribution vary as the current changes. The changing field propagates outward at the speed of light.

There is an associated **magnetic field** (**B**) which propagates outward as well (see [\[link\]](#)). The electric and magnetic fields are closely related and propagate as an electromagnetic wave. This is what happens in broadcast antennae such as those in radio and TV stations.

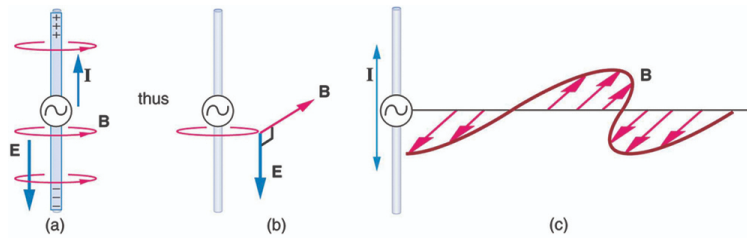
Closer examination of the one complete cycle shown in [\[link\]](#) reveals the periodic nature of the generator-driven charges oscillating up and down in the antenna and the electric field produced. At time $t = 0$, there is the maximum separation of charge, with negative charges at the top and positive charges at the bottom, producing the maximum magnitude of the electric field (or E -field) in the upward direction. One-fourth of a cycle later, there is no charge separation and the field next to the antenna is zero, while the maximum E -field has moved away at speed c .

As the process continues, the charge separation reverses and the field reaches its maximum downward value, returns to zero, and rises to its maximum upward value at the end of one complete cycle. The outgoing wave has an **amplitude** proportional to the maximum separation of charge. Its **wavelength**(λ) is proportional to the period of the oscillation and, hence, is smaller for short periods or high frequencies. (As usual, wavelength and **frequency**(f) are inversely proportional.)

Electric and Magnetic Waves: Moving Together

Following Ampere's law, current in the antenna produces a magnetic field, as shown in [\[link\]](#). The relationship between **E** and **B** is shown at one

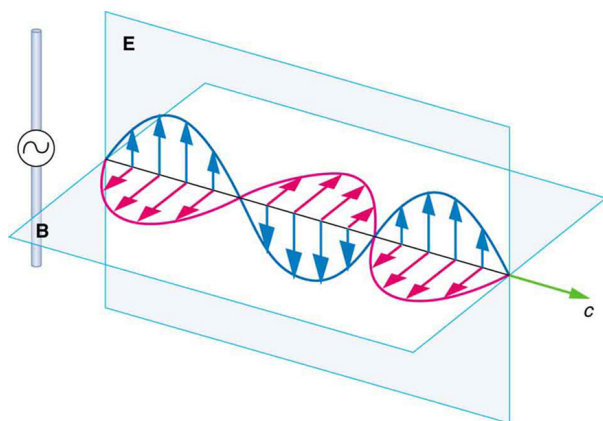
instant in [\[link\]](#) (a). As the current varies, the magnetic field varies in magnitude and direction.



(a) The current in the antenna produces the circular magnetic field lines. The current (I) produces the separation of charge along the wire, which in turn creates the electric field as shown. (b) The electric and magnetic fields ($\mathbf{E}$ and $\mathbf{B}$) near the wire are perpendicular; they are shown here for one point in space. (c) The magnetic field varies with current and propagates away from the antenna at the speed of light.

The magnetic field lines also propagate away from the antenna at the speed of light, forming the other part of the electromagnetic wave, as seen in [\[link\]](#) (b). The magnetic part of the wave has the same period and wavelength as the electric part, since they are both produced by the same movement and separation of charges in the antenna.

The electric and magnetic waves are shown together at one instant in time in [\[link\]](#). The electric and magnetic fields produced by a long straight wire antenna are exactly in phase. Note that they are perpendicular to one another and to the direction of propagation, making this a **transverse wave**.



A part of the electromagnetic wave sent out from the antenna at one instant in time. The electric and magnetic fields (**E** and **B**) are in phase, and they are perpendicular to one another and the direction of propagation. For clarity, the waves are shown only along one direction, but they propagate out in other directions too.

Electromagnetic waves generally propagate out from a source in all directions, sometimes forming a complex radiation pattern. A linear antenna like this one will not radiate parallel to its length, for example. The wave is shown in one direction from the antenna in [\[link\]](#) to illustrate its basic characteristics.

Instead of the AC generator, the antenna can also be driven by an AC circuit. In fact, charges radiate whenever they are accelerated. But while a current in a circuit needs a complete path, an antenna has a varying charge distribution forming a **standing wave**, driven by the AC. The dimensions of the antenna are critical for determining the frequency of the radiated electromagnetic waves. This is a **resonant** phenomenon and when we tune

radios or TV, we vary electrical properties to achieve appropriate resonant conditions in the antenna.

Receiving Electromagnetic Waves

Electromagnetic waves carry energy away from their source, similar to a sound wave carrying energy away from a standing wave on a guitar string. An antenna for receiving EM signals works in reverse. And like antennas that produce EM waves, receiver antennas are specially designed to resonate at particular frequencies.

An incoming electromagnetic wave accelerates electrons in the antenna, setting up a standing wave. If the radio or TV is switched on, electrical components pick up and amplify the signal formed by the accelerating electrons. The signal is then converted to audio and/or video format. Sometimes big receiver dishes are used to focus the signal onto an antenna.

In fact, charges radiate whenever they are accelerated. When designing circuits, we often assume that energy does not quickly escape AC circuits, and mostly this is true. A broadcast antenna is specially designed to enhance the rate of electromagnetic radiation, and shielding is necessary to keep the radiation close to zero. Some familiar phenomena are based on the production of electromagnetic waves by varying currents. Your microwave oven, for example, sends electromagnetic waves, called microwaves, from a concealed antenna that has an oscillating current imposed on it.

Relating E -Field and B -Field Strengths

There is a relationship between the E - and B -field strengths in an electromagnetic wave. This can be understood by again considering the antenna just described. The stronger the E -field created by a separation of charge, the greater the current and, hence, the greater the B -field created.

Since current is directly proportional to voltage (Ohm's law) and voltage is directly proportional to E -field strength, the two should be directly proportional. It can be shown that the magnitudes of the fields do have a constant ratio, equal to the speed of light. That is,

Equation:

$$\frac{E}{B} = c$$

is the ratio of E -field strength to B -field strength in any electromagnetic wave. This is true at all times and at all locations in space. A simple and elegant result.

Example:**Calculating B -Field Strength in an Electromagnetic Wave**

What is the maximum strength of the B -field in an electromagnetic wave that has a maximum E -field strength of 1000 V/m?

Strategy

To find the B -field strength, we rearrange the above equation to solve for B , yielding

Equation:

$$B = \frac{E}{c}.$$

Solution

We are given E , and c is the speed of light. Entering these into the expression for B yields

Equation:

$$B = \frac{1000 \text{ V/m}}{3.00 \times 10^8 \text{ m/s}} = 3.33 \times 10^{-6} \text{ T},$$

Where T stands for Tesla, a measure of magnetic field strength.

Discussion

The B -field strength is less than a tenth of the Earth's admittedly weak magnetic field. This means that a relatively strong electric field of 1000 V/m is accompanied by a relatively weak magnetic field. Note that as this

wave spreads out, say with distance from an antenna, its field strengths become progressively weaker.

The result of this example is consistent with the statement made in the module [Maxwell's Equations: Electromagnetic Waves Predicted and Observed](#) that changing electric fields create relatively weak magnetic fields. They can be detected in electromagnetic waves, however, by taking advantage of the phenomenon of resonance, as Hertz did. A system with the same natural frequency as the electromagnetic wave can be made to oscillate. All radio and TV receivers use this principle to pick up and then amplify weak electromagnetic waves, while rejecting all others not at their resonant frequency.

Note:

Take-Home Experiment: Antennas

For your TV or radio at home, identify the antenna, and sketch its shape. If you don't have cable, you might have an outdoor or indoor TV antenna. Estimate its size. If the TV signal is between 60 and 216 MHz for basic channels, then what is the wavelength of those EM waves?

Try tuning the radio and note the small range of frequencies at which a reasonable signal for that station is received. (This is easier with digital readout.) If you have a car with a radio and extendable antenna, note the quality of reception as the length of the antenna is changed.

Note:

PhET Explorations: Radio Waves and Electromagnetic Fields

Broadcast radio waves from KPhET. Wiggle the transmitter electron manually or have it oscillate automatically. Display the field as a curve or vectors. The strip chart shows the electron positions at the transmitter and at the receiver.

<https://archive.cnx.org/specials/c8dd764c-ae74-11e5-af4c-3375261fa183/radio-waves/#sim-radio-waves>

Section Summary

- Electromagnetic waves are created by oscillating charges (which radiate whenever accelerated) and have the same frequency as the oscillation.
- Since the electric and magnetic fields in most electromagnetic waves are perpendicular to the direction in which the wave moves, it is ordinarily a transverse wave.
- The strengths of the electric and magnetic parts of the wave are related by
Equation:

$$\frac{E}{B} = c,$$

which implies that the magnetic field B is very weak relative to the electric field E .

Conceptual Questions

Exercise:

Problem:

The direction of the electric field shown in each part of [\[link\]](#) is that produced by the charge distribution in the wire. Justify the direction shown in each part, using the Coulomb force law and the definition of $\mathbf{E} = \mathbf{F}/q$, where q is a positive test charge.

Exercise:

Problem:

Is the direction of the magnetic field shown in [\[link\]](#) (a) consistent with the right-hand rule for current (RHR-2) in the direction shown in the figure?

Exercise:

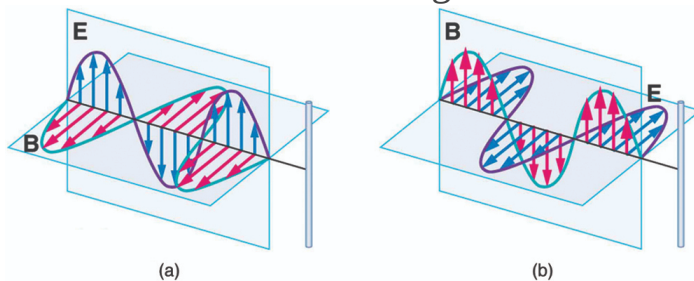
Problem:

Why is the direction of the current shown in each part of [\[link\]](#) opposite to the electric field produced by the wire's charge separation?

Exercise:

Problem:

In which situation shown in [\[link\]](#) will the electromagnetic wave be more successful in inducing a current in the wire? Explain.

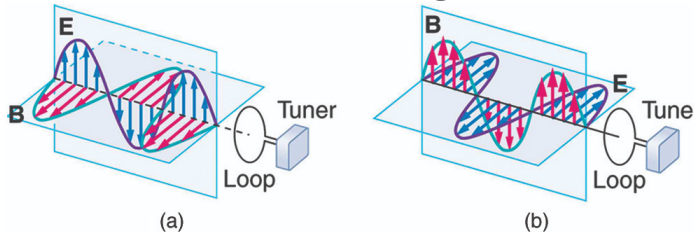


Electromagnetic waves approaching long straight wires.

Exercise:

Problem:

In which situation shown in [\[link\]](#) will the electromagnetic wave be more successful in inducing a current in the loop? Explain.



Electromagnetic waves approaching a wire loop.

Exercise:

Problem:

Should the straight wire antenna of a radio be vertical or horizontal to best receive radio waves broadcast by a vertical transmitter antenna? How should a loop antenna be aligned to best receive the signals? (Note that the direction of the loop that produces the best reception can be used to determine the location of the source. It is used for that purpose in tracking tagged animals in nature studies, for example.)

Exercise:

Problem:

Under what conditions might wires in a DC circuit emit electromagnetic waves?

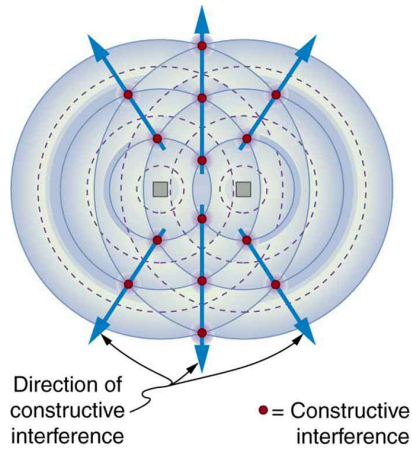
Exercise:

Problem: Give an example of interference of electromagnetic waves.

Exercise:

Problem:

[\[link\]](#) shows the interference pattern of two radio antennas broadcasting the same signal. Explain how this is analogous to the interference pattern for sound produced by two speakers. Could this be used to make a directional antenna system that broadcasts preferentially in certain directions? Explain.



An overhead view
of two radio
broadcast antennas
sending the same
signal, and the
interference pattern
they produce.

Exercise:

Problem: Can an antenna be any length? Explain your answer.

Problems & Exercises

Exercise:

Problem:

What is the maximum electric field strength in an electromagnetic wave that has a maximum magnetic field strength of $5.00 \times 10^{-4} \text{ T}$ (about 10 times the Earth's)?

Solution:

150 kV/m

Exercise:

Problem:

The maximum magnetic field strength of an electromagnetic field is 5×10^{-6} T. Calculate the maximum electric field strength if the wave is traveling in a medium in which the speed of the wave is $0.75c$.

Exercise:

Problem:

Verify the units obtained for magnetic field strength B in [\[link\]](#) (using the equation $B = \frac{E}{c}$) are in fact teslas (T).

Glossary

electric field

a vector quantity (**E**); the lines of electric force per unit charge, moving radially outward from a positive charge and in toward a negative charge

electric field strength

the magnitude of the electric field, denoted E -field

magnetic field

a vector quantity (**B**); can be used to determine the magnetic force on a moving charged particle

magnetic field strength

the magnitude of the magnetic field, denoted B -field

transverse wave

a wave, such as an electromagnetic wave, which oscillates perpendicular to the axis along the line of travel

standing wave

a wave that oscillates in place, with nodes where no motion happens

wavelength

the distance from one peak to the next in a wave

amplitude

the height, or magnitude, of an electromagnetic wave

frequency

the number of complete wave cycles (up-down-up) passing a given point within one second (cycles/second)

resonant

a system that displays enhanced oscillation when subjected to a periodic disturbance of the same frequency as its natural frequency

oscillate

to fluctuate back and forth in a steady beat

The Electromagnetic Spectrum

- List three “rules of thumb” that apply to the different frequencies along the electromagnetic spectrum.
- Explain why the higher the frequency, the shorter the wavelength of an electromagnetic wave.
- Draw a simplified electromagnetic spectrum, indicating the relative positions, frequencies, and spacing of the different types of radiation bands.
- List and explain the different methods by which electromagnetic waves are produced across the spectrum.

In this module we examine how electromagnetic waves are classified into categories such as radio, infrared, ultraviolet, and so on, so that we can understand some of their similarities as well as some of their differences. We will also find that there are many connections with previously discussed topics, such as wavelength and resonance. A brief overview of the production and utilization of electromagnetic waves is found in [\[link\]](#).

Type of EM wave	Production	Applications	Life sciences aspect	Issues
Radio & TV	Accelerating charges	Communications Remote controls	MRI	Requires controls for band use
Microwaves	Accelerating charges & thermal agitation	Communications Ovens Radar	Deep heating	Cell phone use
Infrared	Thermal agitations & electronic transitions	Thermal imaging Heating	Absorbed by atmosphere	Greenhouse effect
Visible light	Thermal agitations & electronic transitions	All pervasive	Photosynthesis Human vision	

Type of EM wave	Production	Applications	Life sciences aspect	Issues
Ultraviolet	Thermal agitations & electronic transitions	Sterilization Cancer control	Vitamin D production	Ozone depletion Cancer causing
X-rays	Inner electronic transitions and fast collisions	Medical Security	Medical diagnosis Cancer therapy	Cancer causing
Gamma rays	Nuclear decay	Nuclear medicineSecurity	Medical diagnosis Cancer therapy	Cancer causing Radiation damage

Electromagnetic Waves

Note:

Connections: Waves

There are many types of waves, such as water waves and even earthquakes. Among the many shared attributes of waves are propagation speed, frequency, and wavelength. These are always related by the expression $v_W = f\lambda$. This module concentrates on EM waves, but other modules contain examples of all of these characteristics for sound waves and submicroscopic particles.

As noted before, an electromagnetic wave has a frequency and a wavelength associated with it and travels at the speed of light, or c . The relationship among these wave characteristics can be described by $v_W = f\lambda$, where v_W is the propagation speed of the wave, f is the frequency, and λ is the wavelength. Here $v_W = c$, so that for all electromagnetic waves,

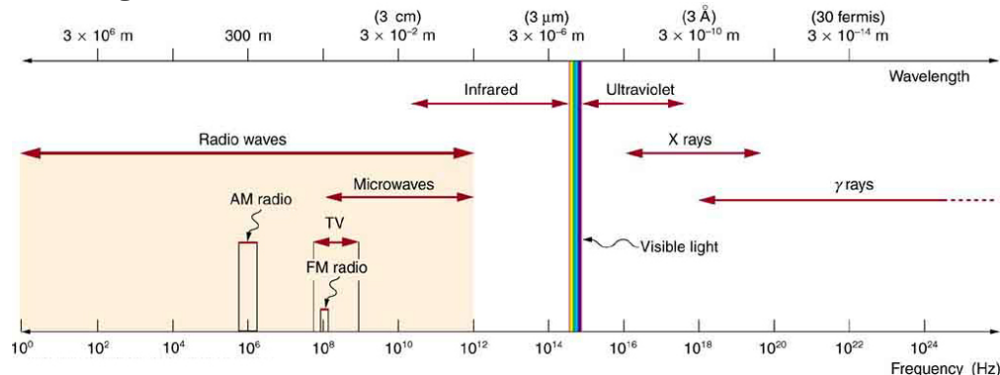
Equation:

$$c = f\lambda.$$

Thus, for all electromagnetic waves, the greater the frequency, the smaller the wavelength.

[\[link\]](#) shows how the various types of electromagnetic waves are categorized according to their wavelengths and frequencies—that is, it shows the electromagnetic spectrum. Many of the

characteristics of the various types of electromagnetic waves are related to their frequencies and wavelengths, as we shall see.



The electromagnetic spectrum, showing the major categories of electromagnetic waves. The range of frequencies and wavelengths is remarkable. The dividing line between some categories is distinct, whereas other categories overlap.

Note:

Electromagnetic Spectrum: Rules of Thumb

Three rules that apply to electromagnetic waves in general are as follows:

- High-frequency electromagnetic waves are more energetic and are more able to penetrate than low-frequency waves.
- High-frequency electromagnetic waves can carry more information per unit time than low-frequency waves.
- The shorter the wavelength of any electromagnetic wave probing a material, the smaller the detail it is possible to resolve.

Note that there are exceptions to these rules of thumb.

Transmission, Reflection, and Absorption

What happens when an electromagnetic wave impinges on a material? If the material is transparent to the particular frequency, then the wave can largely be transmitted. If the material is opaque to the frequency, then the wave can be totally reflected. The wave can also be absorbed by the material, indicating that there is some interaction between the wave and the material, such as the thermal agitation of molecules.

Of course it is possible to have partial transmission, reflection, and absorption. We normally associate these properties with visible light, but they do apply to all electromagnetic waves.

What is not obvious is that something that is transparent to light may be opaque at other frequencies. For example, ordinary glass is transparent to visible light but largely opaque to ultraviolet radiation. Human skin is opaque to visible light—we cannot see through people—but transparent to X-rays.

Radio and TV Waves

The broad category of **radio waves** is defined to contain any electromagnetic wave produced by currents in wires and circuits. Its name derives from their most common use as a carrier of audio information (i.e., radio). The name is applied to electromagnetic waves of similar frequencies regardless of source. Radio waves from outer space, for example, do not come from alien radio stations. They are created by many astronomical phenomena, and their study has revealed much about nature on the largest scales.

There are many uses for radio waves, and so the category is divided into many subcategories, including microwaves and those electromagnetic waves used for AM and FM radio, cellular telephones, and TV.

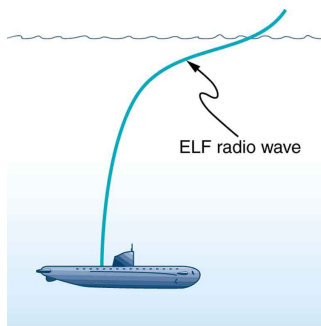
The lowest commonly encountered radio frequencies are produced by high-voltage AC power transmission lines at frequencies of 50 or 60 Hz. (See [\[link\]](#).) These extremely long wavelength electromagnetic waves (about 6000 km!) are one means of energy loss in long-distance power transmission.



This high-voltage traction power line running to Eutingen Railway Substation in Germany radiates electromagnetic waves with very long wavelengths. (credit: Zonk43, Wikimedia Commons)

There is an ongoing controversy regarding potential health hazards associated with exposure to these electromagnetic fields (*E*-fields). Some people suspect that living near such transmission lines may cause a variety of illnesses, including cancer. But demographic data are either inconclusive or simply do not support the hazard theory. Recent reports that have looked at many European and American epidemiological studies have found no increase in risk for cancer due to exposure to *E*-fields.

Extremely low frequency (ELF) radio waves of about 1 kHz are used to communicate with submerged submarines. The ability of radio waves to penetrate salt water is related to their wavelength (much like ultrasound penetrating tissue)—the longer the wavelength, the farther they penetrate. Since salt water is a good conductor, radio waves are strongly absorbed by it, and very long wavelengths are needed to reach a submarine under the surface. (See [\[link\]](#).)

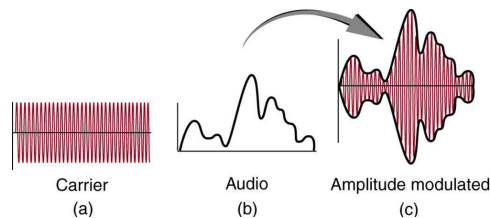


Very long wavelength radio waves are needed to reach this submarine, requiring extremely low frequency signals (ELF). Shorter wavelengths do not penetrate to any significant depth.

AM radio waves are used to carry commercial radio signals in the frequency range from 540 to 1600 kHz. The abbreviation AM stands for **amplitude modulation**, which is the method for placing information on these waves. (See [\[link\]](#).) A **carrier wave** having the basic frequency of the radio station, say 1530 kHz, is varied or modulated in amplitude by an audio signal. The resulting wave has a constant frequency, but a varying amplitude.

A radio receiver tuned to have the same resonant frequency as the carrier wave can pick up the signal, while rejecting the many other frequencies impinging on its antenna. The receiver's

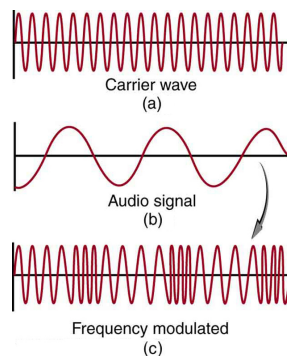
circuitry is designed to respond to variations in amplitude of the carrier wave to replicate the original audio signal. That audio signal is amplified to drive a speaker or perhaps to be recorded.



Amplitude modulation for AM radio. (a) A carrier wave at the station's basic frequency. (b) An audio signal at much lower audible frequencies. (c) The amplitude of the carrier is modulated by the audio signal without changing its basic frequency.

FM Radio Waves

FM radio waves are also used for commercial radio transmission, but in the frequency range of 88 to 108 MHz. FM stands for **frequency modulation**, another method of carrying information. (See [link](#).) Here a carrier wave having the basic frequency of the radio station, perhaps 105.1 MHz, is modulated in frequency by the audio signal, producing a wave of constant amplitude but varying frequency.



Frequency
modulation for

FM radio. (a) A carrier wave at the station's basic frequency. (b) An audio signal at much lower audible frequencies. (c) The frequency of the carrier is modulated by the audio signal without changing its amplitude.

Since audible frequencies range up to 20 kHz (or 0.020 MHz) at most, the frequency of the FM radio wave can vary from the carrier by as much as 0.020 MHz. Thus the carrier frequencies of two different radio stations cannot be closer than 0.020 MHz. An FM receiver is tuned to resonate at the carrier frequency and has circuitry that responds to variations in frequency, reproducing the audio information.

FM radio is inherently less subject to noise from stray radio sources than AM radio. The reason is that amplitudes of waves add. So an AM receiver would interpret noise added onto the amplitude of its carrier wave as part of the information. An FM receiver can be made to reject amplitudes other than that of the basic carrier wave and only look for variations in frequency. It is thus easier to reject noise from FM, since noise produces a variation in amplitude.

Television is also broadcast on electromagnetic waves. Since the waves must carry a great deal of visual as well as audio information, each channel requires a larger range of frequencies than simple radio transmission. TV channels utilize frequencies in the range of 54 to 88 MHz and 174 to 222 MHz. (The entire FM radio band lies between channels 88 MHz and 174 MHz.) These TV channels are called VHF (for **very high frequency**). Other channels called UHF (for **ultra high frequency**) utilize an even higher frequency range of 470 to 1000 MHz.

The TV video signal is AM, while the TV audio is FM. Note that these frequencies are those of free transmission with the user utilizing an old-fashioned roof antenna. Satellite dishes and cable transmission of TV occurs at significantly higher frequencies and is rapidly evolving with the use of the high-definition or HD format.

Example: **Calculating Wavelengths of Radio Waves**

Calculate the wavelengths of a 1530-kHz AM radio signal, a 105.1-MHz FM radio signal, and a 1.90-GHz cell phone signal.

Strategy

The relationship between wavelength and frequency is $c = f\lambda$, where $c = 3.00 \times 10^8$ m/s is the speed of light (the speed of light is only very slightly smaller in air than it is in a vacuum). We can rearrange this equation to find the wavelength for all three frequencies.

Solution

Rearranging gives

Equation:

$$\lambda = \frac{c}{f}.$$

(a) For the $f = 1530$ kHz AM radio signal, then,

Equation:

$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8 \text{ m/s}}{1530 \times 10^3 \text{ cycles/s}} \\ &= 196 \text{ m.}\end{aligned}$$

(b) For the $f = 105.1$ MHz FM radio signal,

Equation:

$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8 \text{ m/s}}{105.1 \times 10^6 \text{ cycles/s}} \\ &= 2.85 \text{ m.}\end{aligned}$$

(c) And for the $f = 1.90$ GHz cell phone,

Equation:

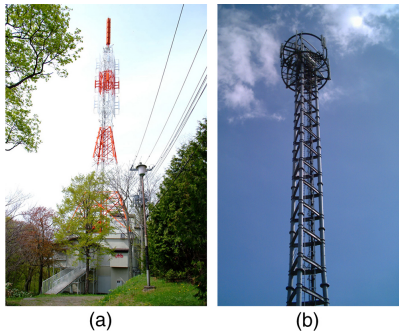
$$\begin{aligned}\lambda &= \frac{3.00 \times 10^8 \text{ m/s}}{1.90 \times 10^9 \text{ cycles/s}} \\ &= 0.158 \text{ m.}\end{aligned}$$

Discussion

These wavelengths are consistent with the spectrum in [\[link\]](#). The wavelengths are also related to other properties of these electromagnetic waves, as we shall see.

The wavelengths found in the preceding example are representative of AM, FM, and cell phones, and account for some of the differences in how they are broadcast and how well they travel. The most efficient length for a linear antenna, such as discussed in [Production of Electromagnetic Waves](#), is $\lambda/2$, half the wavelength of the electromagnetic wave. Thus a very large antenna is needed to efficiently broadcast typical AM radio with its carrier wavelengths on the order of hundreds of meters.

One benefit to these long AM wavelengths is that they can go over and around rather large obstacles (like buildings and hills), just as ocean waves can go around large rocks. FM and TV are best received when there is a line of sight between the broadcast antenna and receiver, and they are often sent from very tall structures. FM, TV, and mobile phone antennas themselves are much smaller than those used for AM, but they are elevated to achieve an unobstructed line of sight. (See [link](#).)



(a) A large tower is used to broadcast TV signals.

The actual antennas are small structures on top of the tower—they are placed at great heights to have a clear line of sight over a large broadcast area. (credit: Ozizo, Wikimedia Commons)

(b) The NTT Dokomo mobile phone tower at Tokorozawa City, Japan.

(credit: tokoroten, Wikimedia Commons)

Radio Wave Interference

Astronomers and astrophysicists collect signals from outer space using electromagnetic waves. A common problem for astrophysicists is the “pollution” from electromagnetic radiation pervading our surroundings from communication systems in general. Even everyday gadgets like our car keys having the facility to lock car doors remotely and being able to turn TVs on and off using remotes involve radio-wave frequencies. In order to prevent interference between all these electromagnetic signals, strict regulations are drawn up for different organizations to utilize different radio frequency bands.

One reason why we are sometimes asked to switch off our mobile phones (operating in the range of 1.9 GHz) on airplanes and in hospitals is that important communications or medical equipment often uses similar radio frequencies and their operation can be affected by frequencies used in the communication devices.

For example, radio waves used in magnetic resonance imaging (MRI) have frequencies on the order of 100 MHz, although this varies significantly depending on the strength of the magnetic field used and the nuclear type being scanned. MRI is an important medical imaging and research tool, producing highly detailed two- and three-dimensional images. Radio waves are broadcast, absorbed, and reemitted in a resonance process that is sensitive to the density of nuclei (usually protons or hydrogen nuclei).

The wavelength of 100-MHz radio waves is 3 m, yet using the sensitivity of the resonant frequency to the magnetic field strength, details smaller than a millimeter can be imaged. This is a good example of an exception to a rule of thumb (in this case, the rubric that details much smaller than the probe's wavelength cannot be detected). The intensity of the radio waves used in MRI presents little or no hazard to human health.

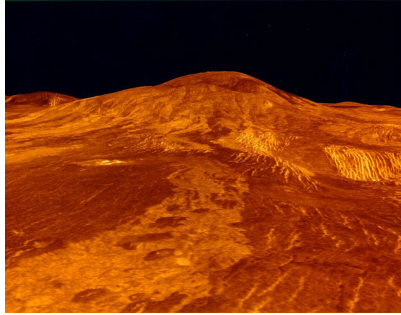
Microwaves

Microwaves are the highest-frequency electromagnetic waves that can be produced by currents in macroscopic circuits and devices. Microwave frequencies range from about 10^9 Hz to the highest practical LC resonance at nearly 10^{12} Hz. Since they have high frequencies, their wavelengths are short compared with those of other radio waves—hence the name “microwave.”

Microwaves can also be produced by atoms and molecules. They are, for example, a component of electromagnetic radiation generated by **thermal agitation**. The thermal motion of atoms and molecules in any object at a temperature above absolute zero causes them to emit and absorb radiation.

Since it is possible to carry more information per unit time on high frequencies, microwaves are quite suitable for communications. Most satellite-transmitted information is carried on microwaves, as are land-based long-distance transmissions. A clear line of sight between transmitter and receiver is needed because of the short wavelengths involved.

Radar is a common application of microwaves that was first developed in World War II. By detecting and timing microwave echoes, radar systems can determine the distance to objects as diverse as clouds and aircraft. A Doppler shift in the radar echo can be used to determine the speed of a car or the intensity of a rainstorm. Sophisticated radar systems are used to map the Earth and other planets, with a resolution limited by wavelength. (See [\[link\]](#).) The shorter the wavelength of any probe, the smaller the detail it is possible to observe.



An image of Sif Mons with lava flows on Venus, based on Magellan synthetic aperture radar data combined with radar altimetry to produce a three-dimensional map of the surface. The Venusian atmosphere is opaque to visible light, but not to the microwaves that were used to create this image.

(credit: NSSDC,
NASA/JPL)

Heating with Microwaves

How does the ubiquitous microwave oven produce microwaves electronically, and why does food absorb them preferentially? Microwaves at a frequency of 2.45 GHz are produced by accelerating electrons. The microwaves are then used to induce an alternating electric field in the oven.

Water and some other constituents of food have a slightly negative charge at one end and a slightly positive charge at one end (called polar molecules). The range of microwave frequencies is specially selected so that the polar molecules, in trying to keep orienting themselves with the electric field, absorb these energies and increase their temperatures—called dielectric heating.

The energy thereby absorbed results in thermal agitation heating food and not the plate, which does not contain water. Hot spots in the food are related to constructive and destructive interference patterns. Rotating antennas and food turntables help spread out the hot spots.

Another use of microwaves for heating is within the human body. Microwaves will penetrate more than shorter wavelengths into tissue and so can accomplish “deep heating” (called

microwave diathermy). This is used for treating muscular pains, spasms, tendonitis, and rheumatoid arthritis.

Note:

Making Connections: Take-Home Experiment—Microwave Ovens

1. Look at the door of a microwave oven. Describe the structure of the door. Why is there a metal grid on the door? How does the size of the holes in the grid compare with the wavelengths of microwaves used in microwave ovens? What is this wavelength?
2. Place a glass of water (about 250 ml) in the microwave and heat it for 30 seconds. Measure the temperature gain (the ΔT). Assuming that the power output of the oven is 1000 W, calculate the efficiency of the heat-transfer process.
3. Remove the rotating turntable or moving plate and place a cup of water in several places along a line parallel with the opening. Heat for 30 seconds and measure the ΔT for each position. Do you see cases of destructive interference?

Microwaves generated by atoms and molecules far away in time and space can be received and detected by electronic circuits. Deep space acts like a blackbody with a 2.7 K temperature, radiating most of its energy in the microwave frequency range. In 1964, Penzias and Wilson detected this radiation and eventually recognized that it was the radiation of the Big Bang's cooled remnants.

Infrared Radiation

The microwave and infrared regions of the electromagnetic spectrum overlap (see [\[link\]](#)).

Infrared radiation is generally produced by thermal motion and the vibration and rotation of atoms and molecules. Electronic transitions in atoms and molecules can also produce infrared radiation.

The range of infrared frequencies extends up to the lower limit of visible light, just below red. In fact, infrared means “below red.” Frequencies at its upper limit are too high to be produced by accelerating electrons in circuits, but small systems, such as atoms and molecules, can vibrate fast enough to produce these waves.

Water molecules rotate and vibrate particularly well at infrared frequencies, emitting and absorbing them so efficiently that the emissivity for skin is $e = 0.97$ in the infrared. Night-vision scopes can detect the infrared emitted by various warm objects, including humans, and convert it to visible light.

We can examine radiant heat transfer from a house by using a camera capable of detecting infrared radiation. Reconnaissance satellites can detect buildings, vehicles, and even individual humans by their infrared emissions, whose power radiation is proportional to the fourth power of the absolute temperature. More mundanely, we use infrared lamps, some of which are called

quartz heaters, to preferentially warm us because we absorb infrared better than our surroundings.

The Sun radiates like a nearly perfect blackbody (that is, it has $e = 1$), with a 6000 K surface temperature. About half of the solar energy arriving at the Earth is in the infrared region, with most of the rest in the visible part of the spectrum, and a relatively small amount in the ultraviolet. On average, 50 percent of the incident solar energy is absorbed by the Earth.

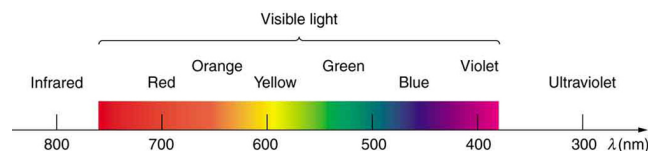
The relatively constant temperature of the Earth is a result of the energy balance between the incoming solar radiation and the energy radiated from the Earth. Most of the infrared radiation emitted from the Earth is absorbed by CO_2 and H_2O in the atmosphere and then radiated back to Earth or into outer space. This radiation back to Earth is known as the greenhouse effect, and it maintains the surface temperature of the Earth about 40°C higher than it would be if there is no absorption. Some scientists think that the increased concentration of CO_2 and other greenhouse gases in the atmosphere, resulting from increases in fossil fuel burning, has increased global average temperatures.

Visible Light

Visible light is the narrow segment of the electromagnetic spectrum to which the normal human eye responds. Visible light is produced by vibrations and rotations of atoms and molecules, as well as by electronic transitions within atoms and molecules. The receivers or detectors of light largely utilize electronic transitions. We say the atoms and molecules are excited when they absorb and relax when they emit through electronic transitions.

[\[link\]](#) shows this part of the spectrum, together with the colors associated with particular pure wavelengths. We usually refer to visible light as having wavelengths of between 400 nm and 750 nm. (The retina of the eye actually responds to the lowest ultraviolet frequencies, but these do not normally reach the retina because they are absorbed by the cornea and lens of the eye.)

Red light has the lowest frequencies and longest wavelengths, while violet has the highest frequencies and shortest wavelengths. Blackbody radiation from the Sun peaks in the visible part of the spectrum but is more intense in the red than in the violet, making the Sun yellowish in appearance.



A small part of the electromagnetic spectrum that includes its visible components. The divisions between infrared, visible, and ultraviolet are not perfectly

distinct, nor are those between the seven rainbow colors.

Living things—plants and animals—have evolved to utilize and respond to parts of the electromagnetic spectrum they are embedded in. Visible light is the most predominant and we enjoy the beauty of nature through visible light. Plants are more selective. Photosynthesis makes use of parts of the visible spectrum to make sugars.

Example:

Integrated Concept Problem: Correcting Vision with Lasers

During laser vision correction, a brief burst of 193-nm ultraviolet light is projected onto the cornea of a patient. It makes a spot 0.80 mm in diameter and evaporates a layer of cornea 0.30 μm thick. Calculate the energy absorbed, assuming the corneal tissue has the same properties as water; it is initially at 34°C. Assume the evaporated tissue leaves at a temperature of 100°C.

Strategy

The energy from the laser light goes toward raising the temperature of the tissue and also toward evaporating it. Thus we have two amounts of heat to add together. Also, we need to find the mass of corneal tissue involved.

Solution

To figure out the heat required to raise the temperature of the tissue to 100°C, we can apply concepts of thermal energy. We know that

Equation:

$$Q = mc\Delta T,$$

where Q is the heat required to raise the temperature, ΔT is the desired change in temperature, m is the mass of tissue to be heated, and c is the specific heat of water equal to 4186 J/kg/K. Without knowing the mass m at this point, we have

Equation:

$$Q = m(4186 \text{ J/kg/K})(100^\circ\text{C} - 34^\circ\text{C}) = m(276,276 \text{ J/kg}) = m(276 \text{ kJ/kg}).$$

The latent heat of vaporization of water is 2256 kJ/kg, so that the energy needed to evaporate mass m is

Equation:

$$Q_v = mL_v = m(2256 \text{ kJ/kg}).$$

To find the mass m , we use the equation $\rho = m/V$, where ρ is the density of the tissue and V is its volume. For this case,

Equation:

$$\begin{aligned}
 m &= \rho V \\
 &= (1000 \text{ kg/m}^3)(\text{area} \times \text{thickness}(\text{m}^3)) \\
 &= (1000 \text{ kg/m}^3)(\pi(0.80 \times 10^{-3} \text{ m})^2/4)(0.30 \times 10^{-6} \text{ m}) \\
 &= 0.151 \times 10^{-9} \text{ kg}.
 \end{aligned}$$

Therefore, the total energy absorbed by the tissue in the eye is the sum of Q and Q_v :

Equation:

$$Q_{\text{tot}} = m(c\Delta T + L_v) = (0.151 \times 10^{-9} \text{ kg})(276 \text{ kJ/kg} + 2256 \text{ kJ/kg}) = 382 \times 10^{-9} \text{ kJ}.$$

Discussion

The lasers used for this eye surgery are excimer lasers, whose light is well absorbed by biological tissue. They evaporate rather than burn the tissue, and can be used for precision work. Most lasers used for this type of eye surgery have an average power rating of about one watt. For our example, if we assume that each laser burst from this pulsed laser lasts for 10 ns, and there are 400 bursts per second, then the average power is $Q_{\text{tot}} \times 400 = 150 \text{ mW}$.

Optics is the study of the behavior of visible light and other forms of electromagnetic waves. Optics falls into two distinct categories. When electromagnetic radiation, such as visible light, interacts with objects that are large compared with its wavelength, its motion can be represented by straight lines like rays. Ray optics is the study of such situations and includes lenses and mirrors.

When electromagnetic radiation interacts with objects about the same size as the wavelength or smaller, its wave nature becomes apparent. For example, observable detail is limited by the wavelength, and so visible light can never detect individual atoms, because they are so much smaller than its wavelength. Physical or wave optics is the study of such situations and includes all wave characteristics.

Note:

Take-Home Experiment: Colors That Match

When you light a match you see largely orange light; when you light a gas stove you see blue light. Why are the colors different? What other colors are present in these?

Ultraviolet Radiation

Ultraviolet means “above violet.” The electromagnetic frequencies of **ultraviolet radiation (UV)** extend upward from violet, the highest-frequency visible light. Ultraviolet is also produced by atomic and molecular motions and electronic transitions. The wavelengths of ultraviolet extend from 400 nm down to about 10 nm at its highest frequencies, which overlap

with the lowest X-ray frequencies. It was recognized as early as 1801 by Johann Ritter that the solar spectrum had an invisible component beyond the violet range.

Solar UV radiation is broadly subdivided into three regions: UV-A (320–400 nm), UV-B (290–320 nm), and UV-C (220–290 nm), ranked from long to shorter wavelengths (from smaller to larger energies). Most UV-B and all UV-C is absorbed by ozone (O₃) molecules in the upper atmosphere. Consequently, 99% of the solar UV radiation reaching the Earth's surface is UV-A.

Human Exposure to UV Radiation

It is largely exposure to UV-B that causes skin cancer. It is estimated that as many as 20% of adults will develop skin cancer over the course of their lifetime. Again, treatment is often successful if caught early. Despite very little UV-B reaching the Earth's surface, there are substantial increases in skin-cancer rates in countries such as Australia, indicating how important it is that UV-B and UV-C continue to be absorbed by the upper atmosphere.

All UV radiation can damage collagen fibers, resulting in an acceleration of the aging process of skin and the formation of wrinkles. Because there is so little UV-B and UV-C reaching the Earth's surface, sunburn is caused by large exposures, and skin cancer from repeated exposure. Some studies indicate a link between overexposure to the Sun when young and melanoma later in life.

The tanning response is a defense mechanism in which the body produces pigments to absorb future exposures in inert skin layers above living cells. Basically UV-B radiation excites DNA molecules, distorting the DNA helix, leading to mutations and the possible formation of cancerous cells.

Repeated exposure to UV-B may also lead to the formation of cataracts in the eyes—a cause of blindness among people living in the equatorial belt where medical treatment is limited. Cataracts, clouding in the eye's lens and a loss of vision, are age related; 60% of those between the ages of 65 and 74 will develop cataracts. However, treatment is easy and successful, as one replaces the lens of the eye with a plastic lens. Prevention is important. Eye protection from UV is more effective with plastic sunglasses than those made of glass.

A major acute effect of extreme UV exposure is the suppression of the immune system, both locally and throughout the body.

Low-intensity ultraviolet is used to sterilize haircutting implements, implying that the energy associated with ultraviolet is deposited in a manner different from lower-frequency electromagnetic waves. (Actually this is true for all electromagnetic waves with frequencies greater than visible light.)

Flash photography is generally not allowed of precious artworks and colored prints because the UV radiation from the flash can cause photo-degradation in the artworks. Often artworks will have an extra-thick layer of glass in front of them, which is especially designed to absorb UV radiation.

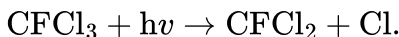
UV Light and the Ozone Layer

If all of the Sun's ultraviolet radiation reached the Earth's surface, there would be extremely grave effects on the biosphere from the severe cell damage it causes. However, the layer of ozone (O₃) in our upper atmosphere (10 to 50 km above the Earth) protects life by absorbing most of the dangerous UV radiation.

Unfortunately, today we are observing a depletion in ozone concentrations in the upper atmosphere. This depletion has led to the formation of an "ozone hole" in the upper atmosphere. The hole is more centered over the southern hemisphere, and changes with the seasons, being largest in the spring. This depletion is attributed to the breakdown of ozone molecules by refrigerant gases called chlorofluorocarbons (CFCs).

The UV radiation helps dissociate the CFC's, releasing highly reactive chlorine (Cl) atoms, which catalyze the destruction of the ozone layer. For example, the reaction of CFC₃ with a photon of light (hν) can be written as:

Equation:



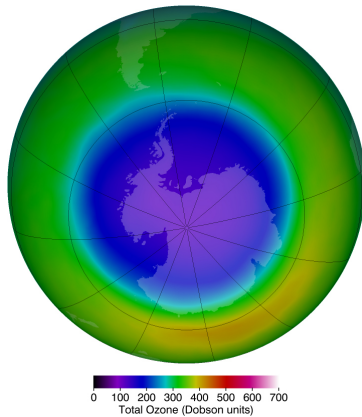
The Cl atom then catalyzes the breakdown of ozone as follows:

Equation:



A single chlorine atom could destroy ozone molecules for up to two years before being transported down to the surface. The CFCs are relatively stable and will contribute to ozone depletion for years to come. CFCs are found in refrigerants, air conditioning systems, foams, and aerosols.

International concern over this problem led to the establishment of the "Montreal Protocol" agreement (1987) to phase out CFC production in most countries. However, developing-country participation is needed if worldwide production and elimination of CFCs is to be achieved. Probably the largest contributor to CFC emissions today is India. But the protocol seems to be working, as there are signs of an ozone recovery. (See [\[link\]](#).)



This map of ozone concentration over Antarctica in October 2011 shows severe depletion suspected to be caused by CFCs.

Less dramatic but more general depletion has been observed over northern latitudes, suggesting the effect is global. With less ozone, more ultraviolet radiation from the Sun reaches the surface, causing more damage. (credit: NASA Ozone Watch)

Benefits of UV Light

Besides the adverse effects of ultraviolet radiation, there are also benefits of exposure in nature and uses in technology. Vitamin D production in the skin (epidermis) results from exposure to UVB radiation, generally from sunlight. A number of studies indicate lack of vitamin D can result in the development of a range of cancers (prostate, breast, colon), so a certain amount of UV exposure is helpful. Lack of vitamin D is also linked to osteoporosis. Exposures (with no sunscreen) of 10 minutes a day to arms, face, and legs might be sufficient to provide the accepted dietary level. However, in the winter time north of about 37° latitude, most UVB gets blocked by the atmosphere.

UV radiation is used in the treatment of infantile jaundice and in some skin conditions. It is also used in sterilizing workspaces and tools, and killing germs in a wide range of applications. It is

also used as an analytical tool to identify substances.

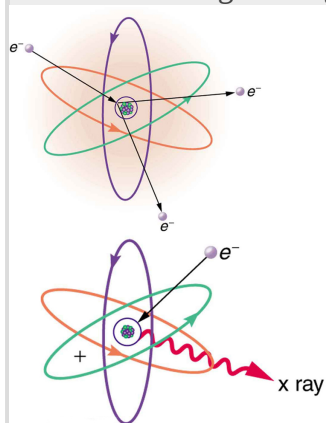
When exposed to ultraviolet, some substances, such as minerals, glow in characteristic visible wavelengths, a process called fluorescence. So-called black lights emit ultraviolet to cause posters and clothing to fluoresce in the visible. Ultraviolet is also used in special microscopes to detect details smaller than those observable with longer-wavelength visible-light microscopes.

Note:

Things Great and Small: A Submicroscopic View of X-Ray Production

X-rays can be created in a high-voltage discharge. They are emitted in the material struck by electrons in the discharge current. There are two mechanisms by which the electrons create X-rays.

The first method is illustrated in [\[link\]](#). An electron is accelerated in an evacuated tube by a high positive voltage. The electron strikes a metal plate (e.g., copper) and produces X-rays. Since this is a high-voltage discharge, the electron gains sufficient energy to ionize the atom.



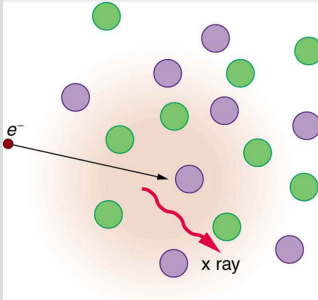
Artist's conception
of an electron
ionizing an atom
followed by the
recapture of an
electron and
emission of an X-
ray. An energetic
electron strikes an
atom and knocks an
electron out of one
of the orbits closest
to the nucleus.

Later, the atom
captures another
electron, and the
energy released by

its fall into a low orbit generates a high-energy EM wave called an X-ray.

In the case shown, an inner-shell electron (one in an orbit relatively close to and tightly bound to the nucleus) is ejected. A short time later, another electron is captured and falls into the orbit in a single great plunge. The energy released by this fall is given to an EM wave known as an X-ray. Since the orbits of the atom are unique to the type of atom, the energy of the X-ray is characteristic of the atom, hence the name characteristic X-ray.

The second method by which an energetic electron creates an X-ray when it strikes a material is illustrated in [\[link\]](#). The electron interacts with charges in the material as it penetrates. These collisions transfer kinetic energy from the electron to the electrons and atoms in the material.



Artist's conception of an electron being slowed by collisions in a material and emitting X-ray radiation. This energetic electron makes numerous collisions with electrons and atoms in a material it penetrates. An accelerated charge radiates EM waves, a second method by which X-rays are created.

A loss of kinetic energy implies an acceleration, in this case decreasing the electron's velocity. Whenever a charge is accelerated, it radiates EM waves. Given the high energy of the electron,

these EM waves can have high energy. We call them X-rays. Since the process is random, a broad spectrum of X-ray energy is emitted that is more characteristic of the electron energy than the type of material the electron encounters. Such EM radiation is called “bremsstrahlung” (German for “braking radiation”).

X-Rays

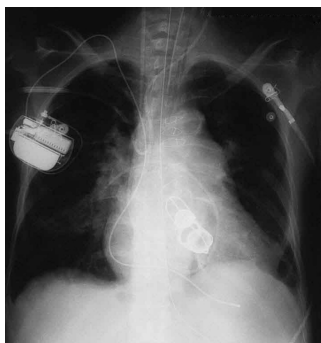
In the 1850s, scientists (such as Faraday) began experimenting with high-voltage electrical discharges in tubes filled with rarefied gases. It was later found that these discharges created an invisible, penetrating form of very high frequency electromagnetic radiation. This radiation was called an **X-ray**, because its identity and nature were unknown.

As described in [Things Great and Small](#), there are two methods by which X-rays are created—both are submicroscopic processes and can be caused by high-voltage discharges. While the low-frequency end of the X-ray range overlaps with the ultraviolet, X-rays extend to much higher frequencies (and energies).

X-rays have adverse effects on living cells similar to those of ultraviolet radiation, and they have the additional liability of being more penetrating, affecting more than the surface layers of cells. Cancer and genetic defects can be induced by exposure to X-rays. Because of their effect on rapidly dividing cells, X-rays can also be used to treat and even cure cancer.

The widest use of X-rays is for imaging objects that are opaque to visible light, such as the human body or aircraft parts. In humans, the risk of cell damage is weighed carefully against the benefit of the diagnostic information obtained. However, questions have risen in recent years as to accidental overexposure of some people during CT scans—a mistake at least in part due to poor monitoring of radiation dose.

The ability of X-rays to penetrate matter depends on density, and so an X-ray image can reveal very detailed density information. [\[link\]](#) shows an example of the simplest type of X-ray image, an X-ray shadow on film. The amount of information in a simple X-ray image is impressive, but more sophisticated techniques, such as CT scans, can reveal three-dimensional information with details smaller than a millimeter.



This shadow X-ray
image shows many
interesting features,
such as artificial
heart valves, a
pacemaker, and the
wires used to close
the sternum.
(credit: P. P. Urone)

The use of X-ray technology in medicine is called radiology—an established and relatively cheap tool in comparison to more sophisticated technologies. Consequently, X-rays are widely available and used extensively in medical diagnostics. During World War I, mobile X-ray units, advocated by Madame Marie Curie, were used to diagnose soldiers.

Because they can have wavelengths less than 0.01 nm, X-rays can be scattered (a process called X-ray diffraction) to detect the shape of molecules and the structure of crystals. X-ray diffraction was crucial to Crick, Watson, and Wilkins in the determination of the shape of the double-helix DNA molecule.

X-rays are also used as a precise tool for trace-metal analysis in X-ray induced fluorescence, in which the energy of the X-ray emissions are related to the specific types of elements and amounts of materials present.

Gamma Rays

Soon after nuclear radioactivity was first detected in 1896, it was found that at least three distinct types of radiation were being emitted. The most penetrating nuclear radiation was called a **gamma ray (γ ray)** (again a name given because its identity and character were unknown), and it was later found to be an extremely high frequency electromagnetic wave.

In fact, γ rays are any electromagnetic radiation emitted by a nucleus. This can be from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the γ -ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation.

Gamma rays have characteristics identical to X-rays of the same frequency—they differ only in source. At higher frequencies, γ rays are more penetrating and more damaging to living tissue. They have many of the same uses as X-rays, including cancer therapy. Gamma radiation from radioactive materials is used in nuclear medicine.

[\[link\]](#) shows a medical image based on γ rays. Food spoilage can be greatly inhibited by exposing it to large doses of γ radiation, thereby obliterating responsible microorganisms. Damage to food cells through irradiation occurs as well, and the long-term hazards of

consuming radiation-preserved food are unknown and controversial for some groups. Both X-ray and γ -ray technologies are also used in scanning luggage at airports.



This is an image of the γ rays emitted by nuclei in a compound that is concentrated in the bones and eliminated through the kidneys. Bone cancer is evidenced by nonuniform concentration in similar

structures.
For example,
some ribs are
darker than
others.
(credit: P. P.
Urone)

Detecting Electromagnetic Waves from Space

A final note on star gazing. The entire electromagnetic spectrum is used by researchers for investigating stars, space, and time. As noted earlier, Penzias and Wilson detected microwaves to identify the background radiation originating from the Big Bang. Radio telescopes such as the Arecibo Radio Telescope in Puerto Rico and Parkes Observatory in Australia were designed to detect radio waves.

Infrared telescopes need to have their detectors cooled by liquid nitrogen to be able to gather useful signals. Since infrared radiation is predominantly from thermal agitation, if the detectors were not cooled, the vibrations of the molecules in the antenna would be stronger than the signal being collected.

The most famous of these infrared sensitive telescopes is the James Clerk Maxwell Telescope in Hawaii. The earliest telescopes, developed in the seventeenth century, were optical telescopes, collecting visible light. Telescopes in the ultraviolet, X-ray, and γ -ray regions are placed outside the atmosphere on satellites orbiting the Earth.

The Hubble Space Telescope (launched in 1990) gathers ultraviolet radiation as well as visible light. In the X-ray region, there is the Chandra X-ray Observatory (launched in 1999), and in the γ -ray region, there is the new Fermi Gamma-ray Space Telescope (launched in 2008—taking the place of the Compton Gamma Ray Observatory, 1991–2000.).

Note:

PhET Explorations: Color Vision

Make a whole rainbow by mixing red, green, and blue light. Change the wavelength of a monochromatic beam or filter white light. View the light as a solid beam, or see the individual photons.

https://phet.colorado.edu/sims/html/color-vision/latest/color-vision_en.html

Section Summary

- The relationship among the speed of propagation, wavelength, and frequency for any wave is given by $v_w = f\lambda$, so that for electromagnetic waves,

Equation:

$$c = f\lambda,$$

where f is the frequency, λ is the wavelength, and c is the speed of light.

- The electromagnetic spectrum is separated into many categories and subcategories, based on the frequency and wavelength, source, and uses of the electromagnetic waves.
- Any electromagnetic wave produced by currents in wires is classified as a radio wave, the lowest frequency electromagnetic waves. Radio waves are divided into many types, depending on their applications, ranging up to microwaves at their highest frequencies.
- Infrared radiation lies below visible light in frequency and is produced by thermal motion and the vibration and rotation of atoms and molecules. Infrared's lower frequencies overlap with the highest-frequency microwaves.
- Visible light is largely produced by electronic transitions in atoms and molecules, and is defined as being detectable by the human eye. Its colors vary with frequency, from red at the lowest to violet at the highest.
- Ultraviolet radiation starts with frequencies just above violet in the visible range and is produced primarily by electronic transitions in atoms and molecules.
- X-rays are created in high-voltage discharges and by electron bombardment of metal targets. Their lowest frequencies overlap the ultraviolet range but extend to much higher values, overlapping at the high end with gamma rays.
- Gamma rays are nuclear in origin and are defined to include the highest-frequency electromagnetic radiation of any type.

Conceptual Questions

Exercise:

Problem:

If you live in a region that has a particular TV station, you can sometimes pick up some of its audio portion on your FM radio receiver. Explain how this is possible. Does it imply that TV audio is broadcast as FM?

Exercise:

Problem:

Explain why people who have the lens of their eye removed because of cataracts are able to see low-frequency ultraviolet.

Exercise:

Problem:

How do fluorescent soap residues make clothing look “brighter and whiter” in outdoor light? Would this be effective in candlelight?

Exercise:

Problem: Give an example of resonance in the reception of electromagnetic waves.

Exercise:

Problem:

Illustrate that the size of details of an object that can be detected with electromagnetic waves is related to their wavelength, by comparing details observable with two different types (for example, radar and visible light or infrared and X-rays).

Exercise:

Problem: Why don't buildings block radio waves as completely as they do visible light?

Exercise:

Problem:

Make a list of some everyday objects and decide whether they are transparent or opaque to each of the types of electromagnetic waves.

Exercise:

Problem:

Your friend says that more patterns and colors can be seen on the wings of birds if viewed in ultraviolet light. Would you agree with your friend? Explain your answer.

Exercise:

Problem:

The rate at which information can be transmitted on an electromagnetic wave is proportional to the frequency of the wave. Is this consistent with the fact that laser telephone transmission at visible frequencies carries far more conversations per optical fiber than conventional electronic transmission in a wire? What is the implication for ELF radio communication with submarines?

Exercise:

Problem: Give an example of energy carried by an electromagnetic wave.

Exercise:

Problem:

In an MRI scan, a higher magnetic field requires higher frequency radio waves to resonate with the nuclear type whose density and location is being imaged. What effect does going to a larger magnetic field have on the most efficient antenna to broadcast those radio waves? Does it favor a smaller or larger antenna?

Exercise:

Problem:

Laser vision correction often uses an excimer laser that produces 193-nm electromagnetic radiation. This wavelength is extremely strongly absorbed by the cornea and ablates it in a manner that reshapes the cornea to correct vision defects. Explain how the strong absorption helps concentrate the energy in a thin layer and thus give greater accuracy in shaping the cornea. Also explain how this strong absorption limits damage to the lens and retina of the eye.

Problems & Exercises**Exercise:****Problem:**

(a) Two microwave frequencies are authorized for use in microwave ovens: 900 and 2560 MHz. Calculate the wavelength of each. (b) Which frequency would produce smaller hot spots in foods due to interference effects?

Solution:

(a) 33.3 cm (900 MHz) 11.7 cm (2560 MHz)

(b) The microwave oven with the smaller wavelength would produce smaller hot spots in foods, corresponding to the one with the frequency 2560 MHz.

Exercise:**Problem:**

(a) Calculate the range of wavelengths for AM radio given its frequency range is 540 to 1600 kHz. (b) Do the same for the FM frequency range of 88.0 to 108 MHz.

Exercise:**Problem:**

A radio station utilizes frequencies between commercial AM and FM. What is the frequency of a 11.12-m-wavelength channel?

Solution:

26.96 MHz

Exercise:**Problem:**

Find the frequency range of visible light, given that it encompasses wavelengths from 380 to 760 nm.

Exercise:**Problem:**

Combing your hair leads to excess electrons on the comb. How fast would you have to move the comb up and down to produce red light?

Solution:

$$5.0 \times 10^{14} \text{ Hz}$$

Exercise:**Problem:**

Electromagnetic radiation having a $15.0 - \mu\text{m}$ wavelength is classified as infrared radiation. What is its frequency?

Exercise:**Problem:**

Approximately what is the smallest detail observable with a microscope that uses ultraviolet light of frequency $1.20 \times 10^{15} \text{ Hz}$?

Solution:**Equation:**

$$\lambda = \frac{c}{f} = \frac{3.00 \times 10^8 \text{ m/s}}{1.20 \times 10^{15} \text{ Hz}} = 2.50 \times 10^{-7} \text{ m}$$

Exercise:**Problem:**

A radar used to detect the presence of aircraft receives a pulse that has reflected off an object $6 \times 10^{-5} \text{ s}$ after it was transmitted. What is the distance from the radar station to the reflecting object?

Exercise:**Problem:**

Some radar systems detect the size and shape of objects such as aircraft and geological terrain. Approximately what is the smallest observable detail utilizing 500-MHz radar?

Solution:

$$0.600 \text{ m}$$

Exercise:

Problem:

Determine the amount of time it takes for X-rays of frequency 3×10^{18} Hz to travel (a) 1 mm and (b) 1 cm.

Exercise:**Problem:**

If you wish to detect details of the size of atoms (about 1×10^{-10} m) with electromagnetic radiation, it must have a wavelength of about this size. (a) What is its frequency? (b) What type of electromagnetic radiation might this be?

Solution:

$$(a) f = \frac{c}{\lambda} = \frac{3.00 \times 10^8 \text{ m/s}}{1 \times 10^{-10} \text{ m}} = 3 \times 10^{18} \text{ Hz}$$

(b) X-rays

Exercise:**Problem:**

If the Sun suddenly turned off, we would not know it until its light stopped coming. How long would that be, given that the Sun is 1.50×10^{11} m away?

Exercise:**Problem:**

Distances in space are often quoted in units of light years, the distance light travels in one year. (a) How many meters is a light year? (b) How many meters is it to Andromeda, the nearest large galaxy, given that it is 2.00×10^6 light years away? (c) The most distant galaxy yet discovered is 12.0×10^9 light years away. How far is this in meters?

Exercise:**Problem:**

A certain 50.0-Hz AC power line radiates an electromagnetic wave having a maximum electric field strength of 13.0 kV/m. (a) What is the wavelength of this very low frequency electromagnetic wave? (b) What is its maximum magnetic field strength?

Solution:

$$(a) 6.00 \times 10^6 \text{ m}$$

$$(b) 4.33 \times 10^{-5} \text{ T}$$

Exercise:

Problem:

During normal beating, the heart creates a maximum 4.00-mV potential across 0.300 m of a person's chest, creating a 1.00-Hz electromagnetic wave. (a) What is the maximum electric field strength created? (b) What is the corresponding maximum magnetic field strength in the electromagnetic wave? (c) What is the wavelength of the electromagnetic wave?

Exercise:**Problem:**

(a) The ideal size (most efficient) for a broadcast antenna with one end on the ground is one-fourth the wavelength ($\lambda/4$) of the electromagnetic radiation being sent out. If a new radio station has such an antenna that is 50.0 m high, what frequency does it broadcast most efficiently? Is this in the AM or FM band? (b) Discuss the analogy of the fundamental resonant mode of an air column closed at one end to the resonance of currents on an antenna that is one-fourth their wavelength.

Solution:

- (a) 1.50×10^6 Hz, AM band
(b) The resonance of currents on an antenna that is $1/4$ their wavelength is analogous to the fundamental resonant mode of an air column closed at one end, since the tube also has a length equal to $1/4$ the wavelength of the fundamental oscillation.

Exercise:**Problem:**

(a) What is the wavelength of 100-MHz radio waves used in an MRI unit? (b) If the frequencies are swept over a ± 1.00 range centered on 100 MHz, what is the range of wavelengths broadcast?

Exercise:**Problem:**

(a) What is the frequency of the 193-nm ultraviolet radiation used in laser eye surgery? (b) Assuming the accuracy with which this EM radiation can ablate the cornea is directly proportional to wavelength, how much more accurate can this UV be than the shortest visible wavelength of light?

Solution:

- (a) 1.55×10^{15} Hz
(b) The shortest wavelength of visible light is 380 nm, so that

Equation:

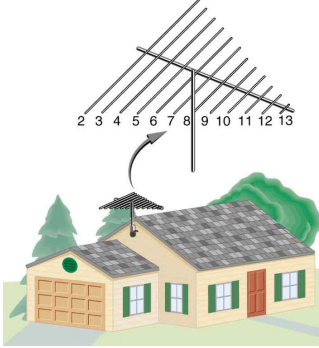
$$\begin{aligned}
 & \frac{\lambda_{\text{visible}}}{\lambda_{\text{UV}}} \\
 &= \frac{380 \text{ nm}}{193 \text{ nm}} \\
 &= 1.97.
 \end{aligned}$$

In other words, the UV radiation is 97% more accurate than the shortest wavelength of visible light, or almost twice as accurate!

Exercise:

Problem:

TV-reception antennas for VHF are constructed with cross wires supported at their centers, as shown in [\[link\]](#). The ideal length for the cross wires is one-half the wavelength to be received, with the more expensive antennas having one for each channel. Suppose you measure the lengths of the wires for particular channels and find them to be 1.94 and 0.753 m long, respectively. What are the frequencies for these channels?



A television reception antenna has cross wires of various lengths to most efficiently receive different wavelengths.

Exercise:

Problem:

Conversations with astronauts on lunar walks had an echo that was used to estimate the distance to the Moon. The sound spoken by the person on Earth was transformed into a radio signal sent to the Moon, and transformed back into sound on a speaker inside the astronaut's space suit. This sound was picked up by the microphone in the space suit (intended for the astronaut's voice) and sent back to Earth as a radio echo of sorts. If the round-trip time was 2.60 s, what was the approximate distance to the Moon, neglecting any delays in the electronics?

Solution:

$$3.90 \times 10^8 \text{ m}$$

Exercise:**Problem:**

Lunar astronauts placed a reflector on the Moon's surface, off which a laser beam is periodically reflected. The distance to the Moon is calculated from the round-trip time. (a) To what accuracy in meters can the distance to the Moon be determined, if this time can be measured to 0.100 ns? (b) What percent accuracy is this, given the average distance to the Moon is $3.84 \times 10^8 \text{ m}$?

Exercise:**Problem:**

Radar is used to determine distances to various objects by measuring the round-trip time for an echo from the object. (a) How far away is the planet Venus if the echo time is 1000 s? (b) What is the echo time for a car 75.0 m from a Highway Police radar unit? (c) How accurately (in nanoseconds) must you be able to measure the echo time to an airplane 12.0 km away to determine its distance within 10.0 m?

Solution:

(a) $1.50 \times 10^{11} \text{ m}$

(b) $0.500 \mu\text{s}$

(c) 66.7 ns

Exercise:**Problem: Integrated Concepts**

(a) Calculate the ratio of the highest to lowest frequencies of electromagnetic waves the eye can see, given the wavelength range of visible light is from 380 to 760 nm. (b) Compare this with the ratio of highest to lowest frequencies the ear can hear.

Exercise:**Problem: Integrated Concepts**

(a) Calculate the rate in watts at which heat transfer through radiation occurs (almost entirely in the infrared) from 1.0 m^2 of the Earth's surface at night. Assume the emissivity is 0.90, the temperature of the Earth is 15°C , and that of outer space is 2.7 K. (b) Compare the intensity of this radiation with that coming to the Earth from the Sun during the day, which averages about 800 W/m^2 , only half of which is absorbed. (c) What is the

maximum magnetic field strength in the outgoing radiation, assuming it is a continuous wave?

Solution:

(a) $-3.5 \times 10^2 \text{ W/m}^2$

(b) 88%

(c) $1.7 \mu\text{T}$

Glossary

electromagnetic spectrum

the full range of wavelengths or frequencies of electromagnetic radiation

radio waves

electromagnetic waves with wavelengths in the range from 1 mm to 100 km; they are produced by currents in wires and circuits and by astronomical phenomena

microwaves

electromagnetic waves with wavelengths in the range from 1 mm to 1 m; they can be produced by currents in macroscopic circuits and devices

thermal agitation

the thermal motion of atoms and molecules in any object at a temperature above absolute zero, which causes them to emit and absorb radiation

radar

a common application of microwaves. Radar can determine the distance to objects as diverse as clouds and aircraft, as well as determine the speed of a car or the intensity of a rainstorm

infrared radiation (IR)

a region of the electromagnetic spectrum with a frequency range that extends from just below the red region of the visible light spectrum up to the microwave region, or from $0.74 \mu\text{m}$ to $300 \mu\text{m}$

ultraviolet radiation (UV)

electromagnetic radiation in the range extending upward in frequency from violet light and overlapping with the lowest X-ray frequencies, with wavelengths from 400 nm down to about 10 nm

visible light

the narrow segment of the electromagnetic spectrum to which the normal human eye responds

amplitude modulation (AM)

a method for placing information on electromagnetic waves by modulating the amplitude of a carrier wave with an audio signal, resulting in a wave with constant frequency but varying amplitude

extremely low frequency (ELF)

electromagnetic radiation with wavelengths usually in the range of 0 to 300 Hz, but also about 1kHz

carrier wave

an electromagnetic wave that carries a signal by modulation of its amplitude or frequency

frequency modulation (FM)

a method of placing information on electromagnetic waves by modulating the frequency of a carrier wave with an audio signal, producing a wave of constant amplitude but varying frequency

TV

video and audio signals broadcast on electromagnetic waves

very high frequency (VHF)

TV channels utilizing frequencies in the two ranges of 54 to 88 MHz and 174 to 222 MHz

ultra-high frequency (UHF)

TV channels in an even higher frequency range than VHF, of 470 to 1000 MHz

X-ray

invisible, penetrating form of very high frequency electromagnetic radiation, overlapping both the ultraviolet range and the γ -ray range

gamma ray

(γ ray); extremely high frequency electromagnetic radiation emitted by the nucleus of an atom, either from natural nuclear decay or induced nuclear processes in nuclear reactors and weapons. The lower end of the γ -ray frequency range overlaps the upper end of the X-ray range, but γ rays can have the highest frequency of any electromagnetic radiation

Energy in Electromagnetic Waves

- Explain how the energy and amplitude of an electromagnetic wave are related.
- Given its power output and the heating area, calculate the intensity of a microwave oven's electromagnetic field, as well as its peak electric and magnetic field strengths

Anyone who has used a microwave oven knows there is energy in **electromagnetic waves**. Sometimes this energy is obvious, such as in the warmth of the summer sun. Other times it is subtle, such as the unfelt energy of gamma rays, which can destroy living cells.

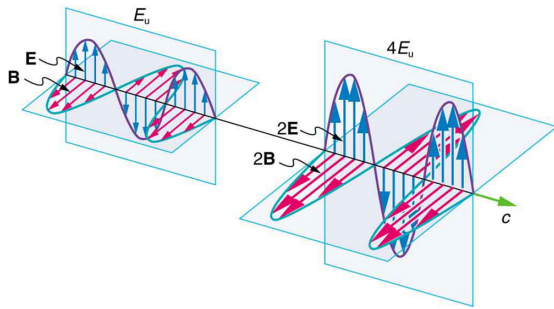
Electromagnetic waves can bring energy into a system by virtue of their **electric and magnetic fields**. These fields can exert forces and move charges in the system and, thus, do work on them. If the frequency of the electromagnetic wave is the same as the natural frequencies of the system (such as microwaves at the resonant frequency of water molecules), the transfer of energy is much more efficient.

Note:

Connections: Waves and Particles

The behavior of electromagnetic radiation clearly exhibits wave characteristics. But we shall find in later modules that at high frequencies, electromagnetic radiation also exhibits particle characteristics. These particle characteristics will be used to explain more of the properties of the electromagnetic spectrum and to introduce the formal study of modern physics.

Another startling discovery of modern physics is that particles, such as electrons and protons, exhibit wave characteristics. This simultaneous sharing of wave and particle properties for all submicroscopic entities is one of the great symmetries in nature.



Energy carried by a wave is proportional to its amplitude squared. With electromagnetic waves, larger E -fields and B -fields exert larger forces and can do more work.

But there is energy in an electromagnetic wave, whether it is absorbed or not. Once created, the fields carry energy away from a source. If absorbed, the field strengths are diminished and anything left travels on. Clearly, the larger the strength of the electric and magnetic fields, the more work they can do and the greater the energy the electromagnetic wave carries.

A wave's energy is proportional to its **amplitude** squared (E^2 or B^2). This is true for waves on guitar strings, for water waves, and for sound waves, where amplitude is proportional to pressure. In electromagnetic waves, the amplitude is the **maximum field strength** of the electric and magnetic fields. (See [\[link\]](#).)

Thus the energy carried and the **intensity** I of an electromagnetic wave is proportional to E^2 and B^2 . In fact, for a continuous sinusoidal electromagnetic wave, the average intensity I_{ave} is given by

Equation:

$$I_{\text{ave}} = \frac{c\epsilon_0 E_0^2}{2},$$

where c is the speed of light, ϵ_0 is the permittivity of free space, and E_0 is the maximum electric field strength; intensity, as always, is power per unit area (here in W/m^2).

The average intensity of an electromagnetic wave I_{ave} can also be expressed in terms of the magnetic field strength by using the relationship $B = E/c$, and the fact that $\epsilon_0 = 1/\mu_0 c^2$, where μ_0 is the permeability of free space. Algebraic manipulation produces the relationship

Equation:

$$I_{\text{ave}} = \frac{cB_0^2}{2\mu_0},$$

where B_0 is the maximum magnetic field strength.

One more expression for I_{ave} in terms of both electric and magnetic field strengths is useful. Substituting the fact that $c \cdot B_0 = E_0$, the previous expression becomes

Equation:

$$I_{\text{ave}} = \frac{E_0 B_0}{2\mu_0}.$$

Whichever of the three preceding equations is most convenient can be used, since they are really just different versions of the same principle: Energy in a wave is related to amplitude squared. Furthermore, since these equations are based on the assumption that the electromagnetic waves are sinusoidal, peak intensity is twice the average; that is, $I_0 = 2I_{\text{ave}}$.

Example:

Calculate Microwave Intensities and Fields

On its highest power setting, a certain microwave oven projects 1.00 kW of microwaves onto a 30.0 by 40.0 cm area. (a) What is the intensity in

W/m²? (b) Calculate the peak electric field strength E_0 in these waves.
(c) What is the peak magnetic field strength B_0 ?

Strategy

In part (a), we can find intensity from its definition as power per unit area. Once the intensity is known, we can use the equations below to find the field strengths asked for in parts (b) and (c).

Solution for (a)

Entering the given power into the definition of intensity, and noting the area is 0.300 by 0.400 m, yields

Equation:

$$I = \frac{P}{A} = \frac{1.00 \text{ kW}}{0.300 \text{ m} \times 0.400 \text{ m}}.$$

Here $I = I_{\text{ave}}$, so that

Equation:

$$I_{\text{ave}} = \frac{1000 \text{ W}}{0.120 \text{ m}^2} = 8.33 \times 10^3 \text{ W/m}^2.$$

Note that the peak intensity is twice the average:

Equation:

$$I_0 = 2I_{\text{ave}} = 1.67 \times 10^4 \text{ W/m}^2.$$

Solution for (b)

To find E_0 , we can rearrange the first equation given above for I_{ave} to give

Equation:

$$E_0 = \left(\frac{2I_{\text{ave}}}{c\epsilon_0} \right)^{1/2}.$$

Entering known values gives

Equation:

$$\begin{aligned}
 E_0 &= \sqrt{\frac{2(8.33 \times 10^3 \text{ W/m}^2)}{(3.00 \times 10^8 \text{ m/s})(8.85 \times 10^{-12} \text{ C}^2/\text{N}\cdot\text{m}^2)}} \\
 &= 2.51 \times 10^3 \text{ V/m}.
 \end{aligned}$$

Solution for (c)

Perhaps the easiest way to find magnetic field strength, now that the electric field strength is known, is to use the relationship given by

Equation:

$$B_0 = \frac{E_0}{c}.$$

Entering known values gives

Equation:

$$\begin{aligned}
 B_0 &= \frac{2.51 \times 10^3 \text{ V/m}}{3.0 \times 10^8 \text{ m/s}} \\
 &= 8.35 \times 10^{-6} \text{ T}.
 \end{aligned}$$

Discussion

As before, a relatively strong electric field is accompanied by a relatively weak magnetic field in an electromagnetic wave, since $B = E/c$, and c is a large number.

Section Summary

- The energy carried by any wave is proportional to its amplitude squared. For electromagnetic waves, this means intensity can be expressed as

Equation:

$$I_{\text{ave}} = \frac{c\varepsilon_0 E_0^2}{2},$$

where I_{ave} is the average intensity in W/m^2 , and E_0 is the maximum electric field strength of a continuous sinusoidal wave.

- This can also be expressed in terms of the maximum magnetic field strength B_0 as

Equation:

$$I_{\text{ave}} = \frac{cB_0^2}{2\mu_0}$$

and in terms of both electric and magnetic fields as

Equation:

$$I_{\text{ave}} = \frac{E_0 B_0}{2\mu_0}.$$

- The three expressions for I_{ave} are all equivalent.

Problems & Exercises

Exercise:

Problem:

What is the intensity of an electromagnetic wave with a peak electric field strength of 125 V/m?

Solution:

Equation:

$$\begin{aligned} I &= \frac{c\varepsilon_0 E_0^2}{2} \\ &= \frac{(3.00 \times 10^8 \text{ m/s})(8.85 \times 10^{-12} \text{ C}^2/\text{N}\cdot\text{m}^2)(125 \text{ V/m})^2}{2} \\ &= 20.7 \text{ W/m}^2 \end{aligned}$$

Exercise:

Problem:

Find the intensity of an electromagnetic wave having a peak magnetic field strength of $4.00 \times 10^{-9} \text{ T}$.

Exercise:**Problem:**

Assume the helium-neon lasers commonly used in student physics laboratories have power outputs of 0.250 mW. (a) If such a laser beam is projected onto a circular spot 1.00 mm in diameter, what is its intensity? (b) Find the peak magnetic field strength. (c) Find the peak electric field strength.

Solution:

$$(a) \ I = \frac{P}{A} = \frac{P}{\pi r^2} = \frac{0.250 \times 10^{-3} \text{ W}}{\pi (0.500 \times 10^{-3} \text{ m})^2} = 318 \text{ W/m}^2$$

$$\begin{aligned} I_{\text{ave}} &= \frac{cB_0^2}{2\mu_0} \Rightarrow B_0 = \left(\frac{2\mu_0 I}{c} \right)^{1/2} \\ (b) &= \left(\frac{2(4\pi \times 10^{-7} \text{ T}\cdot\text{m/A})(318.3 \text{ W/m}^2)}{3.00 \times 10^8 \text{ m/s}} \right)^{1/2} \\ &= 1.63 \times 10^{-6} \text{ T} \end{aligned}$$

$$\begin{aligned} (c) \ E_0 &= cB_0 = (3.00 \times 10^8 \text{ m/s})(1.633 \times 10^{-6} \text{ T}) \\ &= 4.90 \times 10^2 \text{ V/m} \end{aligned}$$

Exercise:

Problem:

An AM radio transmitter broadcasts 50.0 kW of power uniformly in all directions. (a) Assuming all of the radio waves that strike the ground are completely absorbed, and that there is no absorption by the atmosphere or other objects, what is the intensity 30.0 km away? (Hint: Half the power will be spread over the area of a hemisphere.) (b) What is the maximum electric field strength at this distance?

Exercise:**Problem:**

Suppose the maximum safe intensity of microwaves for human exposure is taken to be 1.00 W/m^2 . (a) If a radar unit leaks 10.0 W of microwaves (other than those sent by its antenna) uniformly in all directions, how far away must you be to be exposed to an intensity considered to be safe? Assume that the power spreads uniformly over the area of a sphere with no complications from absorption or reflection. (b) What is the maximum electric field strength at the safe intensity? (Note that early radar units leaked more than modern ones do. This caused identifiable health problems, such as cataracts, for people who worked near them.)

Solution:

(a) 89.2 cm

(b) 27.4 V/m

Exercise:

Problem:

A 2.50-m-diameter university communications satellite dish receives TV signals that have a maximum electric field strength (for one channel) of $7.50 \mu\text{V/m}$. (See [\[link\]](#).) (a) What is the intensity of this wave? (b) What is the power received by the antenna? (c) If the orbiting satellite broadcasts uniformly over an area of $1.50 \times 10^{13} \text{ m}^2$ (a large fraction of North America), how much power does it radiate?



Satellite dishes receive TV signals sent from orbit. Although the signals are quite weak, the receiver can detect them by being tuned to resonate at their frequency.

Exercise:

Problem:

Lasers can be constructed that produce an extremely high intensity electromagnetic wave for a brief time—called pulsed lasers. They are used to ignite nuclear fusion, for example. Such a laser may produce an electromagnetic wave with a maximum electric field strength of $1.00 \times 10^{11} \text{ V/m}$ for a time of 1.00 ns. (a) What is the maximum magnetic field strength in the wave? (b) What is the intensity of the beam? (c) What energy does it deliver on a 1.00-mm^2 area?

Solution:

(a) 333 T

(b) $1.33 \times 10^{19} \text{ W/m}^2$

(c) 13.3 kJ

Exercise:**Problem:**

Show that for a continuous sinusoidal electromagnetic wave, the peak intensity is twice the average intensity ($I_0 = 2I_{\text{ave}}$), using either the fact that $E_0 = \sqrt{2}E_{\text{rms}}$, or $B_0 = \sqrt{2}B_{\text{rms}}$, where rms means average (actually root mean square, a type of average).

Exercise:**Problem:**

Suppose a source of electromagnetic waves radiates uniformly in all directions in empty space where there are no absorption or interference effects. (a) Show that the intensity is inversely proportional to r^2 , the distance from the source squared. (b) Show that the magnitudes of the electric and magnetic fields are inversely proportional to r .

Solution:

$$(a) I = \frac{P}{A} = \frac{P}{4\pi r^2} \propto \frac{1}{r^2}$$

$$(b) I \propto E_0^2, B_0^2 \Rightarrow E_0^2, B_0^2 \propto \frac{1}{r^2} \Rightarrow E_0, B_0 \propto \frac{1}{r}$$

Exercise:

Problem: Integrated Concepts

An LC circuit with a 5.00-pF capacitor oscillates in such a manner as to radiate at a wavelength of 3.30 m. (a) What is the resonant frequency? (b) What inductance is in series with the capacitor?

Exercise:

Problem: Integrated Concepts

What capacitance is needed in series with an $800\text{ } \mu\text{H}$ inductor to form a circuit that radiates a wavelength of 196 m?

Solution:

13.5 pF

Exercise:

Problem: Integrated Concepts

Police radar determines the speed of motor vehicles using the same Doppler-shift technique employed for ultrasound in medical diagnostics. Beats are produced by mixing the double Doppler-shifted echo with the original frequency. If 1.50×10^9 -Hz microwaves are used and a beat frequency of 150 Hz is produced, what is the speed of the vehicle? (Assume the same Doppler-shift formulas are valid with the speed of sound replaced by the speed of light.)

Exercise:

Problem: Integrated Concepts

Assume the mostly infrared radiation from a heat lamp acts like a continuous wave with wavelength $1.50\ \mu\text{m}$. (a) If the lamp's 200-W output is focused on a person's shoulder, over a circular area 25.0 cm in diameter, what is the intensity in W/m^2 ? (b) What is the peak electric field strength? (c) Find the peak magnetic field strength. (d) How long will it take to increase the temperature of the 4.00-kg shoulder by 2.00°C , assuming no other heat transfer and given that its specific heat is $3.47 \times 10^3\ \text{J}/\text{kg}\cdot^\circ\text{C}$?

Solution:

(a) $4.07\ \text{kW}/\text{m}^2$

(b) $1.75\ \text{kV}/\text{m}$

(c) $5.84\ \mu\text{T}$

(d) 2 min 19 s

Exercise:

Problem: Integrated Concepts

On its highest power setting, a microwave oven increases the temperature of 0.400 kg of spaghetti by 45.0°C in 120 s. (a) What was the rate of power absorption by the spaghetti, given that its specific heat is $3.76 \times 10^3\ \text{J}/\text{kg}\cdot^\circ\text{C}$? (b) Find the average intensity of the microwaves, given that they are absorbed over a circular area 20.0 cm in diameter. (c) What is the peak electric field strength of the microwave? (d) What is its peak magnetic field strength?

Exercise:

Problem: Integrated Concepts

Electromagnetic radiation from a 5.00-mW laser is concentrated on a 1.00-mm^2 area. (a) What is the intensity in W/m^2 ? (b) Suppose a 2.00-nC static charge is in the beam. What is the maximum electric

force it experiences? (c) If the static charge moves at 400 m/s, what maximum magnetic force can it feel?

Solution:

(a) $5.00 \times 10^3 \text{ W/m}^2$

(b) $3.88 \times 10^{-6} \text{ N}$

(c) $5.18 \times 10^{-12} \text{ N}$

Exercise:

Problem: Integrated Concepts

A 200-turn flat coil of wire 30.0 cm in diameter acts as an antenna for FM radio at a frequency of 100 MHz. The magnetic field of the incoming electromagnetic wave is perpendicular to the coil and has a maximum strength of $1.00 \times 10^{-12} \text{ T}$. (a) What power is incident on the coil? (b) What average emf is induced in the coil over one-fourth of a cycle? (c) If the radio receiver has an inductance of $2.50 \mu\text{H}$, what capacitance must it have to resonate at 100 MHz?

Exercise:

Problem: Integrated Concepts

If electric and magnetic field strengths vary sinusoidally in time, being zero at $t = 0$, then $E = E_0 \sin 2\pi ft$ and $B = B_0 \sin 2\pi ft$. Let $f = 1.00 \text{ GHz}$ here. (a) When are the field strengths first zero? (b) When do they reach their most negative value? (c) How much time is needed for them to complete one cycle?

Solution:

(a) $t = 0$

(b) $7.50 \times 10^{-10} \text{ s}$

(c) $1.00 \times 10^{-9} \text{ s}$

Exercise:

Problem: Unreasonable Results

A researcher measures the wavelength of a 1.20-GHz electromagnetic wave to be 0.500 m. (a) Calculate the speed at which this wave propagates. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: Unreasonable Results

The peak magnetic field strength in a residential microwave oven is $9.20 \times 10^{-5} \text{ T}$. (a) What is the intensity of the microwave? (b) What is unreasonable about this result? (c) What is wrong about the premise?

Solution:

(a) $1.01 \times 10^6 \text{ W/m}^2$

(b) Much too great for an oven.

(c) The assumed magnetic field is unreasonably large.

Exercise:

Problem: Unreasonable Results

An LC circuit containing a 2.00-H inductor oscillates at such a frequency that it radiates at a 1.00-m wavelength. (a) What is the capacitance of the circuit? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Exercise:

Problem: Unreasonable Results

An LC circuit containing a 1.00-pF capacitor oscillates at such a frequency that it radiates at a 300-nm wavelength. (a) What is the inductance of the circuit? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $2.53 \times 10^{-20} \text{ H}$

(b) L is much too small.

(c) The wavelength is unreasonably small.

Exercise:

Problem: Create Your Own Problem

Consider electromagnetic fields produced by high voltage power lines. Construct a problem in which you calculate the intensity of this electromagnetic radiation in W/m^2 based on the measured magnetic field strength of the radiation in a home near the power lines. Assume these magnetic field strengths are known to average less than a μT . The intensity is small enough that it is difficult to imagine mechanisms for biological damage due to it. Discuss how much energy may be radiating from a section of power line several hundred meters long and compare this to the power likely to be carried by the lines. An idea of how much power this is can be obtained by calculating the approximate current responsible for μT fields at distances of tens of meters.

Exercise:

Problem: Create Your Own Problem

Consider the most recent generation of residential satellite dishes that are a little less than half a meter in diameter. Construct a problem in which you calculate the power received by the dish and the maximum electric field strength of the microwave signals for a single channel

received by the dish. Among the things to be considered are the power broadcast by the satellite and the area over which the power is spread, as well as the area of the receiving dish.

Glossary

maximum field strength

the maximum amplitude an electromagnetic wave can reach, representing the maximum amount of electric force and/or magnetic flux that the wave can exert

intensity

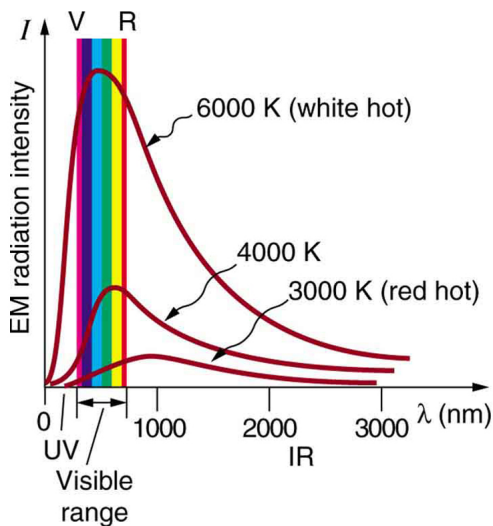
the power of an electric or magnetic field per unit area, for example, Watts per square meter

Quantization of Energy

- Explain Max Planck's contribution to the development of quantum mechanics.
- Explain why atomic spectra indicate quantization.

Planck's Contribution

Energy is quantized in some systems, meaning that the system can have only certain energies and not a continuum of energies, unlike the classical case. This would be like having only certain speeds at which a car can travel because its kinetic energy can have only certain values. We also find that some forms of energy transfer take place with discrete lumps of energy. While most of us are familiar with the quantization of matter into lumps called atoms, molecules, and the like, we are less aware that energy, too, can be quantized. Some of the earliest clues about the necessity of quantum mechanics over classical physics came from the quantization of energy.



Graphs of blackbody radiation (from an ideal radiator) at three different radiator temperatures. The intensity or rate of

radiation emission
increases dramatically
with temperature, and the
peak of the spectrum
shifts toward the visible
and ultraviolet parts of
the spectrum. The shape
of the spectrum cannot be
described with classical
physics.

Where is the quantization of energy observed? Let us begin by considering the emission and absorption of electromagnetic (EM) radiation. The EM spectrum radiated by a hot solid is linked directly to the solid's temperature. (See [\[link\]](#).) An ideal radiator is one that has an emissivity of 1 at all wavelengths and, thus, is jet black. Ideal radiators are therefore called **blackbodies**, and their EM radiation is called **blackbody radiation**. It was discussed that the total intensity of the radiation varies as T^4 , the fourth power of the absolute temperature of the body, and that the peak of the spectrum shifts to shorter wavelengths at higher temperatures. All of this seems quite continuous, but it was the curve of the spectrum of intensity versus wavelength that gave a clue that the energies of the atoms in the solid are quantized. In fact, providing a theoretical explanation for the experimentally measured shape of the spectrum was a mystery at the turn of the century. When this “ultraviolet catastrophe” was eventually solved, the answers led to new technologies such as computers and the sophisticated imaging techniques described in earlier chapters. Once again, physics as an enabling science changed the way we live.

The German physicist Max Planck (1858–1947) used the idea that atoms and molecules in a body act like oscillators to absorb and emit radiation. The energies of the oscillating atoms and molecules had to be quantized to correctly describe the shape of the blackbody spectrum. Planck deduced that the energy of an oscillator having a frequency f is given by

Equation:

$$E = \left(n + \frac{1}{2} \right) hf.$$

Here n is any nonnegative integer (0, 1, 2, 3, ...). The symbol h stands for **Planck's constant**, given by

Equation:

$$h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}.$$

The equation $E = \left(n + \frac{1}{2} \right) hf$ means that an oscillator having a frequency f (emitting and absorbing EM radiation of frequency f) can have its energy increase or decrease only in *discrete* steps of size

Equation:

$$\Delta E = hf.$$

It might be helpful to mention some macroscopic analogies of this quantization of energy phenomena. This is like a pendulum that has a characteristic oscillation frequency but can swing with only certain amplitudes. Quantization of energy also resembles a standing wave on a string that allows only particular harmonics described by integers. It is also similar to going up and down a hill using discrete stair steps rather than being able to move up and down a continuous slope. Your potential energy takes on discrete values as you move from step to step.

Using the quantization of oscillators, Planck was able to correctly describe the experimentally known shape of the blackbody spectrum. This was the first indication that energy is sometimes quantized on a small scale and earned him the Nobel Prize in Physics in 1918. Although Planck's theory comes from observations of a macroscopic object, its analysis is based on atoms and molecules. It was such a revolutionary departure from classical physics that Planck himself was reluctant to accept his own idea that energy states are not continuous. The general acceptance of Planck's energy quantization was greatly enhanced by Einstein's explanation of the photoelectric effect (discussed in the next section), which took energy

quantization a step further. Planck was fully involved in the development of both early quantum mechanics and relativity. He quickly embraced Einstein's special relativity, published in 1905, and in 1906 Planck was the first to suggest the correct formula for relativistic momentum, $p = \gamma mu$.



The German physicist Max Planck had a major influence on the early development of quantum mechanics, being the first to recognize that energy is sometimes quantized. Planck also made important contributions to special relativity and classical physics.
(credit: Library of Congress, Prints and Photographs Division via Wikimedia Commons)

Note that Planck's constant h is a very small number. So for an infrared frequency of 10^{14} Hz being emitted by a blackbody, for example, the difference between energy levels is only $\Delta E = hf = (6.63 \times 10^{-34} \text{ J}\cdot\text{s})(10^{14} \text{ Hz}) = 6.63 \times 10^{-20} \text{ J}$, or about 0.4 eV. This 0.4 eV of energy is significant compared with typical atomic

energies, which are on the order of an electron volt, or thermal energies, which are typically fractions of an electron volt. But on a macroscopic or classical scale, energies are typically on the order of joules. Even if macroscopic energies are quantized, the quantum steps are too small to be noticed. This is an example of the correspondence principle. For a large object, quantum mechanics produces results indistinguishable from those of classical physics.

Atomic Spectra

Now let us turn our attention to the *emission and absorption of EM radiation by gases*. The Sun is the most common example of a body containing gases emitting an EM spectrum that includes visible light. We also see examples in neon signs and candle flames. Studies of emissions of hot gases began more than two centuries ago, and it was soon recognized that these emission spectra contained huge amounts of information. The type of gas and its temperature, for example, could be determined. We now know that these EM emissions come from electrons transitioning between energy levels in individual atoms and molecules; thus, they are called **atomic spectra**. Atomic spectra remain an important analytical tool today. [\[link\]](#) shows an example of an emission spectrum obtained by passing an electric discharge through a material. One of the most important characteristics of these spectra is that they are discrete. By this we mean that only certain wavelengths, and hence frequencies, are emitted. This is called a line spectrum. If frequency and energy are associated as $\Delta E = hf$, the energies of the electrons in the emitting atoms and molecules are quantized. This is discussed in more detail later in this chapter.



Emission spectrum of oxygen. When an electrical discharge is passed through a substance, its atoms and molecules absorb energy, which is reemitted as EM radiation. The discrete nature of these emissions implies that the energy states of the atoms

and molecules are quantized. Such atomic spectra were used as analytical tools for many decades before it was understood why they are quantized. (credit: Teravolt, Wikimedia Commons)

It was a major puzzle that atomic spectra are quantized. Some of the best minds of 19th-century science failed to explain why this might be. Not until the second decade of the 20th century did an answer based on quantum mechanics begin to emerge. Again a macroscopic or classical body of gas was involved in the studies, but the effect, as we shall see, is due to individual atoms and molecules.

Note:

PhET Explorations: Models of the Hydrogen Atom

How did scientists figure out the structure of atoms without looking at them? Try out different models by shooting light at the atom. Check how the prediction of the model matches the experimental results.

<https://archive.cnx.org/specials/d77cc1d0-33e4-11e6-b016-6726afecd2be/hydrogen-atom/#sim-hydrogen-atom>

Section Summary

- The first indication that energy is sometimes quantized came from blackbody radiation, which is the emission of EM radiation by an object with an emissivity of 1.
- Planck recognized that the energy levels of the emitting atoms and molecules were quantized, with only the allowed values of $E = \left(n + \frac{1}{2}\right)hf$, where n is any non-negative integer (0, 1, 2, 3, ...).
- h is Planck's constant, whose value is $h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}$.
- Thus, the oscillatory absorption and emission energies of atoms and molecules in a blackbody could increase or decrease only in steps of

size $\Delta E = hf$ where f is the frequency of the oscillatory nature of the absorption and emission of EM radiation.

- Another indication of energy levels being quantized in atoms and molecules comes from the lines in atomic spectra, which are the EM emissions of individual atoms and molecules.

Conceptual Questions

Exercise:

Problem:

Give an example of a physical entity that is quantized. State specifically what the entity is and what the limits are on its values.

Exercise:

Problem:

Give an example of a physical entity that is not quantized, in that it is continuous and may have a continuous range of values.

Exercise:

Problem:

What aspect of the blackbody spectrum forced Planck to propose quantization of energy levels in its atoms and molecules?

Exercise:

Problem:

If Planck's constant were large, say 10^{34} times greater than it is, we would observe macroscopic entities to be quantized. Describe the motions of a child's swing under such circumstances.

Exercise:

Problem: Why don't we notice quantization in everyday events?

Problems & Exercises

Exercise:

Problem:

A LiBr molecule oscillates with a frequency of 1.7×10^{13} Hz. (a) What is the difference in energy in eV between allowed oscillator states? (b) What is the approximate value of n for a state having an energy of 1.0 eV?

Solution:

(a) 0.070 eV

(b) 14

Exercise:

Problem:

The difference in energy between allowed oscillator states in HBr molecules is 0.330 eV. What is the oscillation frequency of this molecule?

Exercise:

Problem:

A physicist is watching a 15-kg orangutan at a zoo swing lazily in a tire at the end of a rope. He (the physicist) notices that each oscillation takes 3.00 s and hypothesizes that the energy is quantized. (a) What is the difference in energy in joules between allowed oscillator states? (b) What is the value of n for a state where the energy is 5.00 J? (c) Can the quantization be observed?

Solution:

(a) 2.21×10^{-34} J

(b) 2.26×10^{34}

(c) No

Glossary

blackbody

an ideal radiator, which can radiate equally well at all wavelengths

blackbody radiation

the electromagnetic radiation from a blackbody

Planck's constant

$$h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}$$

atomic spectra

the electromagnetic emission from atoms and molecules

The Photoelectric Effect

- Describe a typical photoelectric-effect experiment.
- Determine the maximum kinetic energy of photoelectrons ejected by photons of one energy or wavelength, when given the maximum kinetic energy of photoelectrons for a different photon energy or wavelength.

When light strikes materials, it can eject electrons from them. This is called the **photoelectric effect**, meaning that light (*photo*) produces electricity. One common use of the photoelectric effect is in light meters, such as those that adjust the automatic iris on various types of cameras. In a similar way, another use is in solar cells, as you probably have in your calculator or have seen on a roof top or a roadside sign. These make use of the photoelectric effect to convert light into electricity for running different devices.



The
photoelectric
effect can be
observed by
allowing
light to fall
on the metal
plate in this
evacuated
tube.

Electrons
ejected by
the light are
collected on
the collector
wire and

measured as
a current. A
retarding
voltage
between the
collector
wire and
plate can
then be
adjusted so
as to
determine the
energy of the
ejected
electrons. For
example, if it
is sufficiently
negative, no
electrons will
reach the
wire. (credit:
P.P. Urone)

This effect has been known for more than a century and can be studied using a device such as that shown in [\[link\]](#). This figure shows an evacuated tube with a metal plate and a collector wire that are connected by a variable voltage source, with the collector more negative than the plate. When light (or other EM radiation) strikes the plate in the evacuated tube, it may eject electrons. If the electrons have energy in electron volts (eV) greater than the potential difference between the plate and the wire in volts, some electrons will be collected on the wire. Since the electron energy in eV is qV , where q is the electron charge and V is the potential difference, the electron energy can be measured by adjusting the retarding voltage between the wire and the plate. The voltage that stops the electrons from reaching the wire equals the energy in eV. For example, if -3.00 V barely stops the electrons,

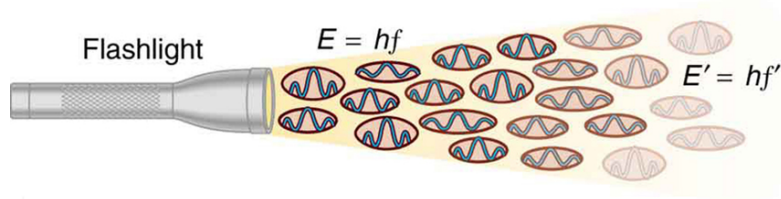
their energy is 3.00 eV. The number of electrons ejected can be determined by measuring the current between the wire and plate. The more light, the more electrons; a little circuitry allows this device to be used as a light meter.

What is really important about the photoelectric effect is what Albert Einstein deduced from it. Einstein realized that there were several characteristics of the photoelectric effect that could be explained only if *EM radiation is itself quantized*: the apparently continuous stream of energy in an EM wave is actually composed of energy quanta called photons. In his explanation of the photoelectric effect, Einstein defined a quantized unit or quantum of EM energy, which we now call a **photon**, with an energy proportional to the frequency of EM radiation. In equation form, the **photon energy** is

Equation:

$$E = hf,$$

where E is the energy of a photon of frequency f and h is Planck's constant. This revolutionary idea looks similar to Planck's quantization of energy states in blackbody oscillators, but it is quite different. It is the quantization of EM radiation itself. EM waves are composed of photons and are not continuous smooth waves as described in previous chapters on optics. Their energy is absorbed and emitted in lumps, not continuously. This is exactly consistent with Planck's quantization of energy levels in blackbody oscillators, since these oscillators increase and decrease their energy in steps of hf by absorbing and emitting photons having $E = hf$. We do not observe this with our eyes, because there are so many photons in common light sources that individual photons go unnoticed. (See [\[link\]](#).) The next section of the text ([Photon Energies and the Electromagnetic Spectrum](#)) is devoted to a discussion of photons and some of their characteristics and implications. For now, we will use the photon concept to explain the photoelectric effect, much as Einstein did.



An EM wave of frequency f is composed of photons, or individual quanta of EM radiation. The energy of each photon is $E = hf$, where h is Planck's constant and f is the frequency of the EM radiation. Higher intensity means more photons per unit area.

The flashlight emits large numbers of photons of many different frequencies, hence others have energy $E' = hf'$, and so on.

The photoelectric effect has the properties discussed below. All these properties are consistent with the idea that individual photons of EM radiation are absorbed by individual electrons in a material, with the electron gaining the photon's energy. Some of these properties are inconsistent with the idea that EM radiation is a simple wave. For simplicity, let us consider what happens with monochromatic EM radiation in which all photons have the same energy hf .

1. If we vary the frequency of the EM radiation falling on a material, we find the following: For a given material, there is a threshold frequency f_0 for the EM radiation below which no electrons are ejected, regardless of intensity. Individual photons interact with individual electrons. Thus if the photon energy is too small to break an electron away, no electrons will be ejected. If EM radiation was a simple wave, sufficient energy could be obtained by increasing the intensity.
2. *Once EM radiation falls on a material, electrons are ejected without delay.* As soon as an individual photon of a sufficiently high frequency is absorbed by an individual electron, the electron is ejected. If the EM

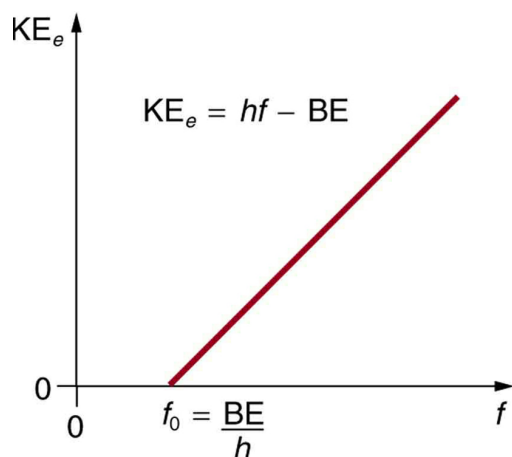
radiation were a simple wave, several minutes would be required for sufficient energy to be deposited to the metal surface to eject an electron.

3. The number of electrons ejected per unit time is proportional to the intensity of the EM radiation and to no other characteristic. High-intensity EM radiation consists of large numbers of photons per unit area, with all photons having the same characteristic energy hf .
4. If we vary the intensity of the EM radiation and measure the energy of ejected electrons, we find the following: *The maximum kinetic energy of ejected electrons is independent of the intensity of the EM radiation.* Since there are so many electrons in a material, it is extremely unlikely that two photons will interact with the same electron at the same time, thereby increasing the energy given it. Instead (as noted in 3 above), increased intensity results in more electrons of the same energy being ejected. If EM radiation were a simple wave, a higher intensity could give more energy, and higher-energy electrons would be ejected.
5. The kinetic energy of an ejected electron equals the photon energy minus the binding energy of the electron in the specific material. An individual photon can give all of its energy to an electron. The photon's energy is partly used to break the electron away from the material. The remainder goes into the ejected electron's kinetic energy. In equation form, this is given by

Equation:

$$KE_e = hf - BE,$$

where KE_e is the maximum kinetic energy of the ejected electron, hf is the photon's energy, and BE is the **binding energy** of the electron to the particular material. (BE is sometimes called the *work function* of the material.) This equation, due to Einstein in 1905, explains the properties of the photoelectric effect quantitatively. An individual photon of EM radiation (it does not come any other way) interacts with an individual electron, supplying enough energy, BE , to break it away, with the remainder going to kinetic energy. The binding energy is $BE = hf_0$, where f_0 is the threshold frequency for the particular material. [\[link\]](#) shows a graph of maximum KE_e versus the frequency of incident EM radiation falling on a particular material.



Photoelectric effect. A graph of the kinetic energy of an ejected electron, KE_e , versus the frequency of EM radiation impinging on a certain material. There is a threshold frequency below which no electrons are ejected, because the individual photon interacting with an individual electron has insufficient energy to break it away. Above the threshold energy, KE_e increases linearly with f , consistent with $KE_e = hf - BE$. The slope of this line is h — the data can be used to determine Planck's constant experimentally. Einstein gave the first successful explanation of such data by proposing

the idea of photons—
quanta of EM radiation.

Einstein's idea that EM radiation is quantized was crucial to the beginnings of quantum mechanics. It is a far more general concept than its explanation of the photoelectric effect might imply. All EM radiation can also be modeled in the form of photons, and the characteristics of EM radiation are entirely consistent with this fact. (As we will see in the next section, many aspects of EM radiation, such as the hazards of ultraviolet (UV) radiation, can be explained *only* by photon properties.) More famous for modern relativity, Einstein planted an important seed for quantum mechanics in 1905, the same year he published his first paper on special relativity. His explanation of the photoelectric effect was the basis for the Nobel Prize awarded to him in 1921. Although his other contributions to theoretical physics were also noted in that award, special and general relativity were not fully recognized in spite of having been partially verified by experiment by 1921. Although hero-worshipped, this great man never received Nobel recognition for his most famous work—relativity.

Example:

Calculating Photon Energy and the Photoelectric Effect: A Violet Light

(a) What is the energy in joules and electron volts of a photon of 420-nm violet light? (b) What is the maximum kinetic energy of electrons ejected from calcium by 420-nm violet light, given that the binding energy (or work function) of electrons for calcium metal is 2.71 eV?

Strategy

To solve part (a), note that the energy of a photon is given by $E = hf$. For part (b), once the energy of the photon is calculated, it is a straightforward application of $KE_e = hf - BE$ to find the ejected electron's maximum kinetic energy, since BE is given.

Solution for (a)

Photon energy is given by

Equation:

$$E = hf$$

Since we are given the wavelength rather than the frequency, we solve the familiar relationship $c = f\lambda$ for the frequency, yielding

Equation:

$$f = \frac{c}{\lambda}.$$

Combining these two equations gives the useful relationship

Equation:

$$E = \frac{hc}{\lambda}.$$

Now substituting known values yields

Equation:

$$E = \frac{(6.63 \times 10^{-34} \text{ J} \cdot \text{s})(3.00 \times 10^8 \text{ m/s})}{420 \times 10^{-9} \text{ m}} = 4.74 \times 10^{-19} \text{ J}.$$

Converting to eV, the energy of the photon is

Equation:

$$E = (4.74 \times 10^{-19} \text{ J}) \frac{1 \text{ eV}}{1.6 \times 10^{-19} \text{ J}} = 2.96 \text{ eV}.$$

Solution for (b)

Finding the kinetic energy of the ejected electron is now a simple application of the equation $\text{KE}_e = hf - \text{BE}$. Substituting the photon energy and binding energy yields

Equation:

$$\text{KE}_e = hf - \text{BE} = 2.96 \text{ eV} - 2.71 \text{ eV} = 0.246 \text{ eV}.$$

Discussion

The energy of this 420-nm photon of violet light is a tiny fraction of a joule, and so it is no wonder that a single photon would be difficult for us to sense directly—humans are more attuned to energies on the order of joules. But looking at the energy in electron volts, we can see that this photon has enough energy to affect atoms and molecules. A DNA molecule can be broken with about 1 eV of energy, for example, and typical atomic and molecular energies are on the order of eV, so that the UV photon in this example could have biological effects. The ejected electron (called a *photoelectron*) has a rather low energy, and it would not travel far, except in a vacuum. The electron would be stopped by a retarding potential of but 0.26 eV. In fact, if the photon wavelength were longer and its energy less than 2.71 eV, then the formula would give a negative kinetic energy, an impossibility. This simply means that the 420-nm photons with their 2.96-eV energy are not much above the frequency threshold. You can show for yourself that the threshold wavelength is 459 nm (blue light). This means that if calcium metal is used in a light meter, the meter will be insensitive to wavelengths longer than those of blue light. Such a light meter would be completely insensitive to red light, for example.

Note:**PhET Explorations: Photoelectric Effect**

See how light knocks electrons off a metal target, and recreate the experiment that spawned the field of quantum mechanics.

<https://archive.cnx.org/specials/cf1152da-eae8-11e5-b874-f779884a9994/photoelectric-effect/#sim-photoelectric-effect>

Section Summary

- The photoelectric effect is the process in which EM radiation ejects electrons from a material.
- Einstein proposed photons to be quanta of EM radiation having energy $E = hf$, where f is the frequency of the radiation.

- All EM radiation is composed of photons. As Einstein explained, all characteristics of the photoelectric effect are due to the interaction of individual photons with individual electrons.
- The maximum kinetic energy KE_e of ejected electrons (photoelectrons) is given by $KE_e = hf - BE$, where hf is the photon energy and BE is the binding energy (or work function) of the electron to the particular material.

Conceptual Questions

Exercise:

Problem:

Is visible light the only type of EM radiation that can cause the photoelectric effect?

Exercise:

Problem:

Which aspects of the photoelectric effect cannot be explained without photons? Which can be explained without photons? Are the latter inconsistent with the existence of photons?

Exercise:

Problem:

Is the photoelectric effect a direct consequence of the wave character of EM radiation or of the particle character of EM radiation? Explain briefly.

Exercise:

Problem:

Insulators (nonmetals) have a higher BE than metals, and it is more difficult for photons to eject electrons from insulators. Discuss how this relates to the free charges in metals that make them good conductors.

Exercise:**Problem:**

If you pick up and shake a piece of metal that has electrons in it free to move as a current, no electrons fall out. Yet if you heat the metal, electrons can be boiled off. Explain both of these facts as they relate to the amount and distribution of energy involved with shaking the object as compared with heating it.

Problems & Exercises**Exercise:****Problem:**

What is the longest-wavelength EM radiation that can eject a photoelectron from silver, given that the binding energy is 4.73 eV? Is this in the visible range?

Solution:

263 nm

Exercise:**Problem:**

Find the longest-wavelength photon that can eject an electron from potassium, given that the binding energy is 2.24 eV. Is this visible EM radiation?

Exercise:**Problem:**

What is the binding energy in eV of electrons in magnesium, if the longest-wavelength photon that can eject electrons is 337 nm?

Solution:

3.69 eV

Exercise:

Problem:

Calculate the binding energy in eV of electrons in aluminum, if the longest-wavelength photon that can eject them is 304 nm.

Exercise:

Problem:

What is the maximum kinetic energy in eV of electrons ejected from sodium metal by 450-nm EM radiation, given that the binding energy is 2.28 eV?

Solution:

0.483 eV

Exercise:

Problem:

UV radiation having a wavelength of 120 nm falls on gold metal, to which electrons are bound by 4.82 eV. What is the maximum kinetic energy of the ejected photoelectrons?

Exercise:

Problem:

Violet light of wavelength 400 nm ejects electrons with a maximum kinetic energy of 0.860 eV from sodium metal. What is the binding energy of electrons to sodium metal?

Solution:

2.25 eV

Exercise:

Problem:

UV radiation having a 300-nm wavelength falls on uranium metal, ejecting 0.500-eV electrons. What is the binding energy of electrons to uranium metal?

Exercise:**Problem:**

What is the wavelength of EM radiation that ejects 2.00-eV electrons from calcium metal, given that the binding energy is 2.71 eV? What type of EM radiation is this?

Solution:

- (a) 264 nm
- (b) Ultraviolet

Exercise:**Problem:**

Find the wavelength of photons that eject 0.100-eV electrons from potassium, given that the binding energy is 2.24 eV. Are these photons visible?

Exercise:**Problem:**

What is the maximum velocity of electrons ejected from a material by 80-nm photons, if they are bound to the material by 4.73 eV?

Solution:

$$1.95 \times 10^6 \text{ m/s}$$

Exercise:

Problem:

Photoelectrons from a material with a binding energy of 2.71 eV are ejected by 420-nm photons. Once ejected, how long does it take these electrons to travel 2.50 cm to a detection device?

Exercise:**Problem:**

A laser with a power output of 2.00 mW at a wavelength of 400 nm is projected onto calcium metal. (a) How many electrons per second are ejected? (b) What power is carried away by the electrons, given that the binding energy is 2.71 eV?

Solution:

(a) $4.02 \times 10^{15} \text{ /s}$

(b) 0.256 mW

Exercise:**Problem:**

(a) Calculate the number of photoelectrons per second ejected from a 1.00-mm^2 area of sodium metal by 500-nm EM radiation having an intensity of 1.30 kW/m^2 (the intensity of sunlight above the Earth's atmosphere). (b) Given that the binding energy is 2.28 eV, what power is carried away by the electrons? (c) The electrons carry away less power than brought in by the photons. Where does the other power go? How can it be recovered?

Exercise:**Problem: Unreasonable Results**

Red light having a wavelength of 700 nm is projected onto magnesium metal to which electrons are bound by 3.68 eV. (a) Use $KE_e = hf - BE$ to calculate the kinetic energy of the ejected electrons.

(b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) -1.90 eV

(b) Negative kinetic energy

(c) That the electrons would be knocked free.

Exercise:

Problem: Unreasonable Results

(a) What is the binding energy of electrons to a material from which 4.00-eV electrons are ejected by 400-nm EM radiation? (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Glossary

photoelectric effect

the phenomenon whereby some materials eject electrons when light is shined on them

photon

a quantum, or particle, of electromagnetic radiation

photon energy

the amount of energy a photon has; $E = hf$

binding energy

also called the *work function*; the amount of energy necessary to eject an electron from a material

Photon Energies and the Electromagnetic Spectrum

- Explain the relationship between the energy of a photon in joules or electron volts and its wavelength or frequency.
- Calculate the number of photons per second emitted by a monochromatic source of specific wavelength and power.

Ionizing Radiation

A photon is a quantum of EM radiation. Its energy is given by $E = hf$ and is related to the frequency f and wavelength λ of the radiation by

Equation:

$$E = hf = \frac{hc}{\lambda} (\text{energy of a photon}),$$

where E is the energy of a single photon and c is the speed of light. When working with small systems, energy in eV is often useful. Note that Planck's constant in these units is

Equation:

$$h = 4.14 \times 10^{-15} \text{ eV} \cdot \text{s}.$$

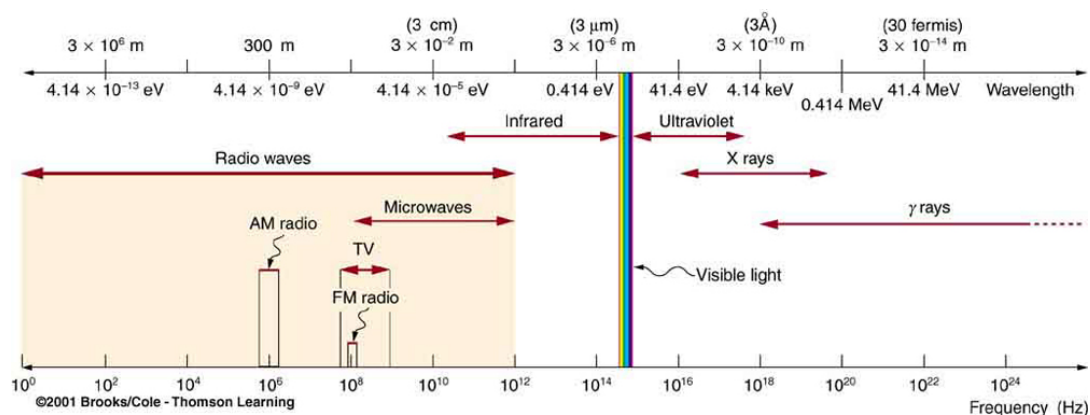
Since many wavelengths are stated in nanometers (nm), it is also useful to know that

Equation:

$$hc = 1240 \text{ eV} \cdot \text{nm}.$$

These will make many calculations a little easier.

All EM radiation is composed of photons. [\[link\]](#) shows various divisions of the EM spectrum plotted against wavelength, frequency, and photon energy. Previously in this book, photon characteristics were alluded to in the discussion of some of the characteristics of UV, x rays, and γ rays, the first of which start with frequencies just above violet in the visible spectrum. It was noted that these types of EM radiation have characteristics much different than visible light. We can now see that such properties arise because photon energy is larger at high frequencies.



The EM spectrum, showing major categories as a function of photon energy in eV, as well as wavelength and frequency. Certain characteristics of EM radiation are directly attributable to photon energy alone.

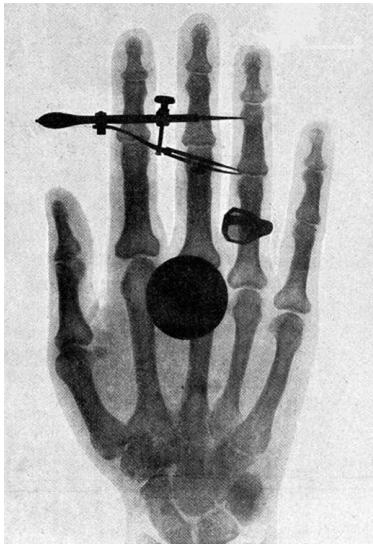
Rotational energies of molecules	10^{-5} eV
Vibrational energies of molecules	0.1 eV
Energy between outer electron shells in atoms	1 eV
Binding energy of a weakly bound molecule	1 eV
Energy of red light	2 eV
Binding energy of a tightly bound molecule	10 eV
Energy to ionize atom or molecule	10 to 1000 eV

Representative Energies for Submicroscopic Effects (Order of Magnitude Only)

Photons act as individual quanta and interact with individual electrons, atoms, molecules, and so on. The energy a photon carries is, thus, crucial to the effects it has. [\[link\]](#) lists representative submicroscopic energies in eV. When we compare photon energies from the EM spectrum in [\[link\]](#) with energies in the table, we can see how effects vary with the type of EM radiation.

Gamma rays, a form of nuclear and cosmic EM radiation, can have the highest frequencies and, hence, the highest photon energies in the EM spectrum. For example, a γ -ray photon with $f = 10^{21}$ Hz has an energy $E = hf = 6.63 \times 10^{-13} \text{ J} = 4.14 \text{ MeV}$. This is sufficient energy to ionize thousands of atoms and molecules, since only 10 to 1000 eV are needed per ionization. In fact, γ rays are one type of **ionizing radiation**, as are x rays and UV, because they produce ionization in materials that absorb

them. Because so much ionization can be produced, a single γ -ray photon can cause significant damage to biological tissue, killing cells or damaging their ability to properly reproduce. When cell reproduction is disrupted, the result can be cancer, one of the known effects of exposure to ionizing radiation. Since cancer cells are rapidly reproducing, they are exceptionally sensitive to the disruption produced by ionizing radiation. This means that ionizing radiation has positive uses in cancer treatment as well as risks in producing cancer.



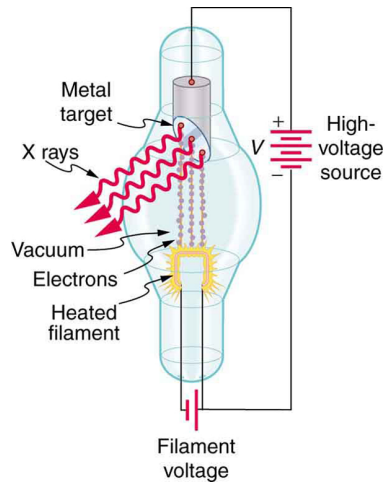
One of the first x-ray images, taken by Röntgen himself. The hand belongs to Bertha Röntgen, his wife.
(credit: Wilhelm Conrad Röntgen, via Wikimedia Commons)

High photon energy also enables γ rays to penetrate materials, since a collision with a single atom or molecule is unlikely to absorb all the γ ray's energy. This can make γ rays useful as a probe, and they are sometimes used in medical imaging. **x rays**, as you can see in [\[link\]](#), overlap with the low-frequency end of the γ ray range. Since x rays have energies of keV and up, individual x-ray photons also can produce large amounts of ionization. At lower photon energies, x rays are not as penetrating as γ rays and are slightly less hazardous. X rays are ideal for medical imaging, their most common use, and a fact that was recognized immediately upon their discovery in 1895 by the German physicist W. C. Roentgen (1845–1923). (See [\[link\]](#).) Within one year of their discovery, x rays (for a time called Roentgen rays) were used for medical diagnostics. Roentgen received the 1901 Nobel Prize for the discovery of x rays.

Note:

Connections: Conservation of Energy

Once again, we find that conservation of energy allows us to consider the initial and final forms that energy takes, without having to make detailed calculations of the intermediate steps. [\[link\]](#) is solved by considering only the initial and final forms of energy.



X rays are produced when energetic electrons strike the copper anode of this cathode ray tube (CRT). Electrons (shown here as separate particles) interact individually with the material they strike, sometimes producing photons of EM radiation.

While γ rays originate in nuclear decay, x rays are produced by the process shown in [\[link\]](#). Electrons ejected by thermal agitation from a hot filament in a vacuum tube are accelerated through a high voltage, gaining kinetic energy from the electrical potential energy. When they strike the anode, the electrons convert their kinetic energy to a variety of forms, including thermal energy. But since an accelerated charge radiates EM waves, and since the electrons act individually, photons are also produced. Some of these x-ray photons obtain the kinetic energy of the electron. The accelerated electrons originate at the cathode, so such a tube is called a cathode ray tube (CRT), and various versions of them are found in older TV and computer screens as well as in x-ray machines.

Example:

X-ray Photon Energy and X-ray Tube Voltage

Find the maximum energy in eV of an x-ray photon produced by electrons accelerated through a potential difference of 50.0 kV in a CRT like the one in [\[link\]](#).

Strategy

Electrons can give all of their kinetic energy to a single photon when they strike the anode of a CRT. (This is something like the photoelectric effect in reverse.) The kinetic energy of the electron comes from electrical potential energy. Thus we can simply equate the maximum photon energy to the electrical potential energy—that is, $hf = qV$. (We do not have to calculate each step from beginning to end if we know that all of the starting energy qV is converted to the final form hf .)

Solution

The maximum photon energy is $hf = qV$, where q is the charge of the electron and V is the accelerating voltage. Thus,

Equation:

$$hf = (1.60 \times 10^{-19} \text{ C})(50.0 \times 10^3 \text{ V}).$$

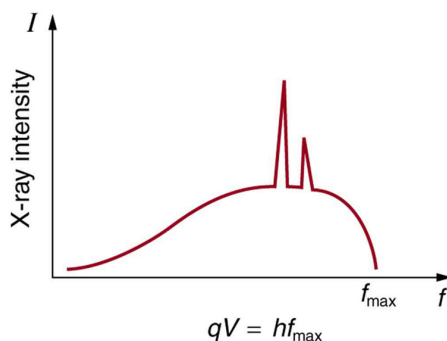
From the definition of the electron volt, we know $1 \text{ eV} = 1.60 \times 10^{-19} \text{ J}$, where $1 \text{ J} = 1 \text{ C} \cdot \text{V}$. Gathering factors and converting energy to eV yields

Equation:

$$hf = (50.0 \times 10^3)(1.60 \times 10^{-19} \text{ C} \cdot \text{V}) \left(\frac{1 \text{ eV}}{1.60 \times 10^{-19} \text{ C} \cdot \text{V}} \right) = (50.0 \times 10^3)(1 \text{ eV}) = 50.0 \text{ keV}.$$

Discussion

This example produces a result that can be applied to many similar situations. If you accelerate a single elementary charge, like that of an electron, through a potential given in volts, then its energy in eV has the same numerical value. Thus a 50.0-kV potential generates 50.0 keV electrons, which in turn can produce photons with a maximum energy of 50 keV. Similarly, a 100-kV potential in an x-ray tube can generate up to 100-keV x-ray photons. Many x-ray tubes have adjustable voltages so that various energy x rays with differing energies, and therefore differing abilities to penetrate, can be generated.



X-ray spectrum obtained when energetic electrons strike a material. The smooth part of the spectrum is bremsstrahlung, while the peaks are characteristic of the anode material. Both are atomic processes that produce energetic

photons known as x-ray
photons.

[\[link\]](#) shows the spectrum of x rays obtained from an x-ray tube. There are two distinct features to the spectrum. First, the smooth distribution results from electrons being decelerated in the anode material. A curve like this is obtained by detecting many photons, and it is apparent that the maximum energy is unlikely. This decelerating process produces radiation that is called **bremsstrahlung** (German for *braking radiation*). The second feature is the existence of sharp peaks in the spectrum; these are called **characteristic x rays**, since they are characteristic of the anode material. Characteristic x rays come from atomic excitations unique to a given type of anode material. They are akin to lines in atomic spectra, implying the energy levels of atoms are quantized. Phenomena such as discrete atomic spectra and characteristic x rays are explored further in [Atomic Physics](#).

Ultraviolet radiation (approximately 4 eV to 300 eV) overlaps with the low end of the energy range of x rays, but UV is typically lower in energy. UV comes from the de-excitation of atoms that may be part of a hot solid or gas. These atoms can be given energy that they later release as UV by numerous processes, including electric discharge, nuclear explosion, thermal agitation, and exposure to x rays. A UV photon has sufficient energy to ionize atoms and molecules, which makes its effects different from those of visible light. UV thus has some of the same biological effects as γ rays and x rays. For example, it can cause skin cancer and is used as a sterilizer. The major difference is that several UV photons are required to disrupt cell reproduction or kill a bacterium, whereas single γ -ray and X-ray photons can do the same damage. But since UV does have the energy to alter molecules, it can do what visible light cannot. One of the beneficial aspects of UV is that it triggers the production of vitamin D in the skin, whereas visible light has insufficient energy per photon to alter the molecules that trigger this production. Infantile jaundice is treated by exposing the baby to UV (with eye protection), called phototherapy, the beneficial effects of which are thought to be related to its ability to help prevent the buildup of potentially toxic bilirubin in the blood.

Example:

Photon Energy and Effects for UV

Short-wavelength UV is sometimes called vacuum UV, because it is strongly absorbed by air and must be studied in a vacuum. Calculate the photon energy in eV for 100-nm vacuum UV, and estimate the number of molecules it could ionize or break apart.

Strategy

Using the equation $E = hf$ and appropriate constants, we can find the photon energy and compare it with energy information in [\[link\]](#).

Solution

The energy of a photon is given by

Equation:

$$E = hf = \frac{hc}{\lambda}.$$

Using $hc = 1240 \text{ eV} \cdot \text{nm}$, we find that

Equation:

$$E = \frac{hc}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{100 \text{ nm}} = 12.4 \text{ eV}.$$

Discussion

According to [\[link\]](#), this photon energy might be able to ionize an atom or molecule, and it is about what is needed to break up a tightly bound molecule, since they are bound by approximately 10 eV. This photon energy could destroy about a dozen weakly bound molecules. Because of its high photon energy, UV disrupts atoms and molecules it interacts with. One good consequence is that all but the longest-wavelength UV is strongly absorbed and is easily blocked by sunglasses. In fact, most of the Sun's UV is absorbed by a thin layer of ozone in the upper atmosphere, protecting sensitive organisms on Earth. Damage to our ozone layer by the addition of such chemicals as CFC's has reduced this protection for us.

Visible Light

The range of photon energies for **visible light** from red to violet is 1.63 to 3.26 eV, respectively (left for this chapter's Problems and Exercises to verify). These energies are on the order of those between outer electron shells in atoms and molecules. This means that these photons can be absorbed by atoms and molecules. A *single* photon can actually stimulate the retina, for example, by altering a receptor molecule that then triggers a nerve impulse. Photons can be absorbed or emitted only by atoms and molecules that have precisely the correct quantized energy step to do so. For example, if a red photon of frequency f encounters a molecule that has an energy step, ΔE , equal to hf , then the photon can be absorbed. Violet flowers absorb red and reflect violet; this implies there is no energy step between levels in the receptor molecule equal to the violet photon's energy, but there is an energy step for the red.

There are some noticeable differences in the characteristics of light between the two ends of the visible spectrum that are due to photon energies. Red light has insufficient photon energy to expose most black-and-white film, and it is thus used to illuminate darkrooms where such film is developed. Since violet light has a higher photon energy, dyes that absorb violet tend to fade more quickly than those that do not. (See [\[link\]](#).) Take a look at some faded color posters in a storefront some time, and you will notice that the blues and violets are the last to fade. This is because other dyes, such as red and green dyes, absorb blue and violet photons, the higher energies of which break up their weakly bound molecules. (Complex molecules such as those in dyes and DNA tend to be weakly bound.) Blue and violet dyes reflect those colors and, therefore, do not absorb these more energetic photons, thus suffering less molecular damage.



Why do the reds, yellows,
and greens fade before
the blues and violets
when exposed to the Sun,
as with this poster? The
answer is related to
photon energy. (credit:
Deb Collins, Flickr)

Transparent materials, such as some glasses, do not absorb any visible light, because there is no energy step in the atoms or molecules that could absorb the light. Since individual photons interact with individual atoms, it is nearly impossible to have two photons absorbed simultaneously to reach a large energy step. Because of its lower photon energy, visible light can sometimes pass through many kilometers of a substance, while higher frequencies like UV, x ray, and γ rays are absorbed, because they have sufficient photon energy to ionize the material.

Example:

How Many Photons per Second Does a Typical Light Bulb Produce?

Assuming that 10.0% of a 100-W light bulb's energy output is in the visible range (typical for incandescent bulbs) with an average wavelength of 580 nm, calculate the number of visible photons emitted per second.

Strategy

Power is energy per unit time, and so if we can find the energy per photon, we can determine the number of photons per second. This will best be done in joules, since power is given in watts, which are joules per second.

Solution

The power in visible light production is 10.0% of 100 W, or 10.0 J/s. The energy of the average visible photon is found by substituting the given average wavelength into the formula

Equation:

$$E = \frac{hc}{\lambda}.$$

This produces

Equation:

$$E = \frac{(6.63 \times 10^{-34} \text{ J} \cdot \text{s})(3.00 \times 10^8 \text{ m/s})}{580 \times 10^{-9} \text{ m}} = 3.43 \times 10^{-19} \text{ J}.$$

The number of visible photons per second is thus

Equation:

$$\text{photon/s} = \frac{10.0 \text{ J/s}}{3.43 \times 10^{-19} \text{ J/photon}} = 2.92 \times 10^{19} \text{ photon/s}.$$

Discussion

This incredible number of photons per second is verification that individual photons are insignificant in ordinary human experience. It is also a verification of the correspondence principle—on the macroscopic scale, quantization becomes essentially continuous or classical. Finally, there are so many photons emitted by a 100-W lightbulb that it can be seen by the unaided eye many kilometers away.

Lower-Energy Photons

Infrared radiation (IR) has even lower photon energies than visible light and cannot significantly alter atoms and molecules. IR can be absorbed and emitted by atoms and molecules, particularly between closely spaced states. IR is extremely strongly absorbed by water, for example, because water molecules have many states separated by energies on the order of 10^{-5} eV to 10^{-2} eV, well within the IR and microwave energy ranges. This is why in the IR range, skin is almost jet black, with an emissivity near 1—there are many states in water molecules in the skin that can absorb a large range of IR photon energies. Not all molecules have this property. Air, for example, is nearly transparent to many IR frequencies.

Microwaves are the highest frequencies that can be produced by electronic circuits, although they are also produced naturally. Thus microwaves are similar to IR but do not extend to as high frequencies. There are states in water and other molecules that have the same frequency and energy as microwaves, typically about 10^{-5} eV. This is one reason why food absorbs microwaves more strongly than many other materials, making microwave ovens an efficient way of putting energy directly into food.

Photon energies for both IR and microwaves are so low that huge numbers of photons are involved in any significant energy transfer by IR or microwaves (such as warming yourself with a heat lamp or cooking pizza in the microwave). Visible light, IR, microwaves, and all lower frequencies cannot produce ionization with single photons and do not ordinarily have the hazards of higher frequencies. When visible, IR, or microwave radiation is hazardous, such as the inducement of cataracts by microwaves, the hazard is due to huge numbers of photons acting together (not to an accumulation of photons, such as sterilization by weak UV). The negative effects of visible, IR, or microwave radiation can be thermal effects, which could be produced by any heat source. But one difference is that at very high intensity, strong electric and magnetic fields can be produced by photons acting together. Such electromagnetic fields (EMF) can actually ionize materials.

Note:**Misconception Alert: High-Voltage Power Lines**

Although some people think that living near high-voltage power lines is hazardous to one's health, ongoing studies of the transient field effects produced by these lines show their strengths to be insufficient to cause damage. Demographic studies also fail to show significant correlation of ill effects with high-voltage power lines. The American Physical Society issued a report over 10 years ago on power-line fields, which concluded that the scientific literature and reviews of panels show no consistent, significant link between cancer and power-line fields. They also felt that the "diversion of resources to eliminate a threat which has no persuasive scientific basis is disturbing."

It is virtually impossible to detect individual photons having frequencies below microwave frequencies, because of their low photon energy. But the photons are there. A continuous EM wave can be modeled as photons. At low frequencies, EM waves are generally treated as time- and position-varying electric and magnetic fields with no discernible quantization. This is another example of the correspondence principle in situations involving huge numbers of photons.

Note:**PhET Explorations: Color Vision**

Make a whole rainbow by mixing red, green, and blue light. Change the wavelength of a monochromatic beam or filter white light. View the light as a solid beam, or see the individual photons.

https://phet.colorado.edu/sims/html/color-vision/latest/color-vision_en.html

Section Summary

- Photon energy is responsible for many characteristics of EM radiation, being particularly noticeable at high frequencies.
- Photons have both wave and particle characteristics.

Conceptual Questions**Exercise:**

Problem: Why are UV, x rays, and γ rays called ionizing radiation?

Exercise:

Problem:

How can treating food with ionizing radiation help keep it from spoiling? UV is not very penetrating. What else could be used?

Exercise:

Problem:

Some television tubes are CRTs. They use an approximately 30-kV accelerating potential to send electrons to the screen, where the electrons stimulate phosphors to emit the light that forms the pictures we watch. Would you expect x rays also to be created?

Exercise:**Problem:**

Tanning salons use “safe” UV with a longer wavelength than some of the UV in sunlight. This “safe” UV has enough photon energy to trigger the tanning mechanism. Is it likely to be able to cause cell damage and induce cancer with prolonged exposure?

Exercise:**Problem:**

Your pupils dilate when visible light intensity is reduced. Does wearing sunglasses that lack UV blockers increase or decrease the UV hazard to your eyes? Explain.

Exercise:**Problem:**

One could feel heat transfer in the form of infrared radiation from a large nuclear bomb detonated in the atmosphere 75 km from you. However, none of the profusely emitted x rays or γ rays reaches you. Explain.

Exercise:

Problem: Can a single microwave photon cause cell damage? Explain.

Exercise:**Problem:**

In an x-ray tube, the maximum photon energy is given by $hf = qV$. Would it be technically more correct to say $hf = qV + BE$, where BE is the binding energy of electrons in the target anode? Why isn't the energy stated the latter way?

Problems & Exercises**Exercise:****Problem:**

What is the energy in joules and eV of a photon in a radio wave from an AM station that has a 1530-kHz broadcast frequency?

Solution:

$$6.34 \times 10^{-9} \text{ eV}, 1.01 \times 10^{-27} \text{ J}$$

Exercise:

Problem:

(a) Find the energy in joules and eV of photons in radio waves from an FM station that has a 90.0-MHz broadcast frequency. (b) What does this imply about the number of photons per second that the radio station must broadcast?

Exercise:

Problem: Calculate the frequency in hertz of a 1.00-MeV γ -ray photon.

Solution:

$$2.42 \times 10^{20} \text{ Hz}$$

Exercise:**Problem:**

(a) What is the wavelength of a 1.00-eV photon? (b) Find its frequency in hertz. (c) Identify the type of EM radiation.

Exercise:**Problem:**

Do the unit conversions necessary to show that $hc = 1240 \text{ eV} \cdot \text{nm}$, as stated in the text.

Solution:**Equation:**

$$\begin{aligned} hc &= (6.62607 \times 10^{-34} \text{ J} \cdot \text{s}) (2.99792 \times 10^8 \text{ m/s}) \left(\frac{10^9 \text{ nm}}{1 \text{ m}} \right) \left(\frac{1.00000 \text{ eV}}{1.60218 \times 10^{-19} \text{ J}} \right) \\ &= 1239.84 \text{ eV} \cdot \text{nm} \\ &\approx 1240 \text{ eV} \cdot \text{nm} \end{aligned}$$

Exercise:**Problem:**

Confirm the statement in the text that the range of photon energies for visible light is 1.63 to 3.26 eV, given that the range of visible wavelengths is 380 to 760 nm.

Exercise:**Problem:**

(a) Calculate the energy in eV of an IR photon of frequency $2.00 \times 10^{13} \text{ Hz}$. (b) How many of these photons would need to be absorbed simultaneously by a tightly bound molecule to break it apart? (c) What is the energy in eV of a γ ray of frequency $3.00 \times 10^{20} \text{ Hz}$? (d) How many tightly bound molecules could a single such γ ray break apart?

Solution:

(a) 0.0829 eV

- (b) 121
- (c) 1.24 MeV
- (d) 1.24×10^5

Exercise:

Problem: Prove that, to three-digit accuracy, $h = 4.14 \times 10^{-15} \text{ eV} \cdot \text{s}$, as stated in the text.

Exercise:

Problem:

(a) What is the maximum energy in eV of photons produced in a CRT using a 25.0-kV accelerating potential, such as a color TV? (b) What is their frequency?

Solution:

- (a) $25.0 \times 10^3 \text{ eV}$
- (b) $6.04 \times 10^{18} \text{ Hz}$

Exercise:

Problem:

What is the accelerating voltage of an x-ray tube that produces x rays with a shortest wavelength of 0.0103 nm?

Exercise:

Problem:

(a) What is the ratio of power outputs by two microwave ovens having frequencies of 950 and 2560 MHz, if they emit the same number of photons per second? (b) What is the ratio of photons per second if they have the same power output?

Solution:

- (a) 2.69
- (b) 0.371

Exercise:

Problem:

How many photons per second are emitted by the antenna of a microwave oven, if its power output is 1.00 kW at a frequency of 2560 MHz?

Exercise:

Problem:

Some satellites use nuclear power. (a) If such a satellite emits a 1.00-W flux of γ rays having an average energy of 0.500 MeV, how many are emitted per second? (b) These γ rays affect other satellites. How far away must another satellite be to only receive one γ ray per second per square meter?

Solution:

(a) 1.25×10^{13} photons/s

(b) 997 km

Exercise:**Problem:**

(a) If the power output of a 650-kHz radio station is 50.0 kW, how many photons per second are produced? (b) If the radio waves are broadcast uniformly in all directions, find the number of photons per second per square meter at a distance of 100 km. Assume no reflection from the ground or absorption by the air.

Exercise:**Problem:**

How many x-ray photons per second are created by an x-ray tube that produces a flux of x rays having a power of 1.00 W? Assume the average energy per photon is 75.0 keV.

Solution:

8.33×10^{13} photons/s

Exercise:**Problem:**

(a) How far away must you be from a 650-kHz radio station with power 50.0 kW for there to be only one photon per second per square meter? Assume no reflections or absorption, as if you were in deep outer space. (b) Discuss the implications for detecting intelligent life in other solar systems by detecting their radio broadcasts.

Exercise:**Problem:**

Assuming that 10.0% of a 100-W light bulb's energy output is in the visible range (typical for incandescent bulbs) with an average wavelength of 580 nm, and that the photons spread out uniformly and are not absorbed by the atmosphere, how far away would you be if 500 photons per second enter the 3.00-mm diameter pupil of your eye? (This number easily stimulates the retina.)

Solution:

181 km

Exercise:

Problem:Construct Your Own Problem

Consider a laser pen. Construct a problem in which you calculate the number of photons per second emitted by the pen. Among the things to be considered are the laser pen's wavelength and power output. Your instructor may also wish for you to determine the minimum diffraction spreading in the beam and the number of photons per square centimeter the pen can project at some large distance. In this latter case, you will also need to consider the output size of the laser beam, the distance to the object being illuminated, and any absorption or scattering along the way.

Glossary

gamma ray

also γ -ray; highest-energy photon in the EM spectrum

ionizing radiation

radiation that ionizes materials that absorb it

x ray

EM photon between γ -ray and UV in energy

bremsstrahlung

German for *braking radiation*; produced when electrons are decelerated

characteristic x rays

x rays whose energy depends on the material they were produced in

ultraviolet radiation

UV; ionizing photons slightly more energetic than violet light

visible light

the range of photon energies the human eye can detect

infrared radiation

photons with energies slightly less than red light

microwaves

photons with wavelengths on the order of a micron (μm)

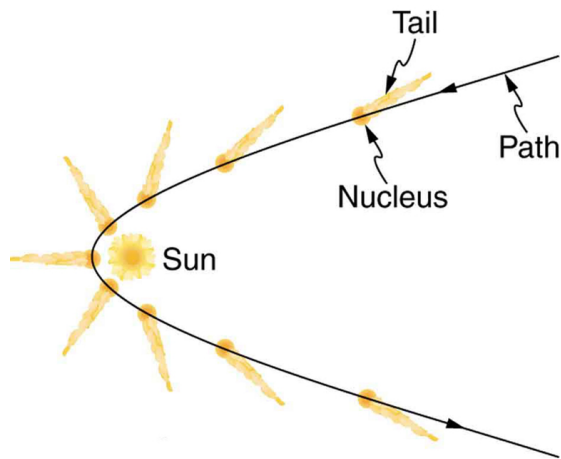
Photon Momentum

- Relate the linear momentum of a photon to its energy or wavelength, and apply linear momentum conservation to simple processes involving the emission, absorption, or reflection of photons.
- Account qualitatively for the increase of photon wavelength that is observed, and explain the significance of the Compton wavelength.

Measuring Photon Momentum

The quantum of EM radiation we call a **photon** has properties analogous to those of particles we can see, such as grains of sand. A photon interacts as a unit in collisions or when absorbed, rather than as an extensive wave.

Massive quanta, like electrons, also act like macroscopic particles—something we expect, because they are the smallest units of matter. Particles carry momentum as well as energy. Despite photons having no mass, there has long been evidence that EM radiation carries momentum. (Maxwell and others who studied EM waves predicted that they would carry momentum.) It is now a well-established fact that photons *do* have momentum. In fact, photon momentum is suggested by the photoelectric effect, where photons knock electrons out of a substance. [\[link\]](#) shows macroscopic evidence of photon momentum.



The tails of the Hale-Bopp comet point away from the Sun, evidence that light has momentum. Dust emanating from the body of the comet forms this tail. Particles of dust are pushed away from the Sun by light reflecting from them. The blue ionized gas tail is also produced by photons interacting with atoms in the comet material. (credit: Geoff Chester, U.S. Navy, via Wikimedia Commons)

[\[link\]](#) shows a comet with two prominent tails. What most people do not know about the tails is that they always point *away* from the Sun rather than trailing behind the comet (like the tail of Bo Peep's sheep). Comet tails are composed of gases and dust evaporated from the body of the comet and ionized gas. The dust particles recoil away from the Sun when photons scatter from them. Evidently, photons carry momentum in the direction of their motion (away from the Sun), and some of this momentum is transferred to dust particles in collisions. Gas atoms and molecules in the blue tail are most affected by other particles of radiation, such as protons and electrons emanating from the Sun, rather than by the momentum of photons.

Note:

Connections: Conservation of Momentum

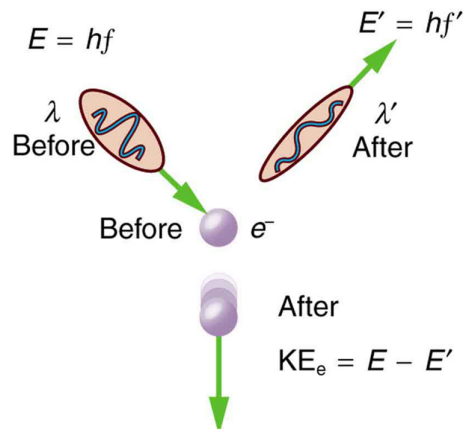
Not only is momentum conserved in all realms of physics, but all types of particles are found to have momentum. We expect particles with mass to have momentum, but now we see that massless particles including photons also carry momentum.

Momentum is conserved in quantum mechanics just as it is in relativity and classical physics. Some of the earliest direct experimental evidence of this came from scattering of x-ray photons by electrons in substances, named Compton scattering after the American physicist Arthur H. Compton (1892–1962). Around 1923, Compton observed that x rays scattered from materials had a decreased energy and correctly analyzed this as being due to the scattering of photons from electrons. This phenomenon could be handled as a collision between two particles—a photon and an electron at rest in the material. Energy and momentum are conserved in the collision. (See [\[link\]](#)) He won a Nobel Prize in 1929 for the discovery of this scattering, now called the **Compton effect**, because it helped prove that **photon momentum** is given by

Equation:

$$p = \frac{h}{\lambda},$$

where h is Planck's constant and λ is the photon wavelength. (Note that relativistic momentum given as $p = \gamma mu$ is valid only for particles having mass.)



The Compton effect is the name given to the scattering of a photon by an electron. Energy and momentum are conserved, resulting in a reduction of both for the scattered photon. Studying this effect, Compton verified that photons have momentum.

We can see that photon momentum is small, since $p = h/\lambda$ and h is very small. It is for this reason that we do not ordinarily observe photon

momentum. Our mirrors do not recoil when light reflects from them (except perhaps in cartoons). Compton saw the effects of photon momentum because he was observing x rays, which have a small wavelength and a relatively large momentum, interacting with the lightest of particles, the electron.

Example:

Electron and Photon Momentum Compared

(a) Calculate the momentum of a visible photon that has a wavelength of 500 nm. (b) Find the velocity of an electron having the same momentum. (c) What is the energy of the electron, and how does it compare with the energy of the photon?

Strategy

Finding the photon momentum is a straightforward application of its definition: $p = \frac{h}{\lambda}$. If we find the photon momentum is small, then we can assume that an electron with the same momentum will be nonrelativistic, making it easy to find its velocity and kinetic energy from the classical formulas.

Solution for (a)

Photon momentum is given by the equation:

Equation:

$$p = \frac{h}{\lambda}.$$

Entering the given photon wavelength yields

Equation:

$$p = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{500 \times 10^{-9} \text{ m}} = 1.33 \times 10^{-27} \text{ kg} \cdot \text{m/s}.$$

Solution for (b)

Since this momentum is indeed small, we will use the classical expression $p = mv$ to find the velocity of an electron with this momentum. Solving for v and using the known value for the mass of an electron gives

Equation:

$$v = \frac{p}{m} = \frac{1.33 \times 10^{-27} \text{ kg} \cdot \text{m/s}}{9.11 \times 10^{-31} \text{ kg}} = 1460 \text{ m/s} \approx 1460 \text{ m/s}.$$

Solution for (c)

The electron has kinetic energy, which is classically given by

Equation:

$$\text{KE}_e = \frac{1}{2}mv^2.$$

Thus,

Equation:

$$\text{KE}_e = \frac{1}{2}(9.11 \times 10^{-31} \text{ kg})(1455 \text{ m/s})^2 = 9.64 \times 10^{-25} \text{ J}.$$

Converting this to eV by multiplying by $(1 \text{ eV})/(1.602 \times 10^{-19} \text{ J})$ yields

Equation:

$$\text{KE}_e = 6.02 \times 10^{-6} \text{ eV}.$$

The photon energy E is

Equation:

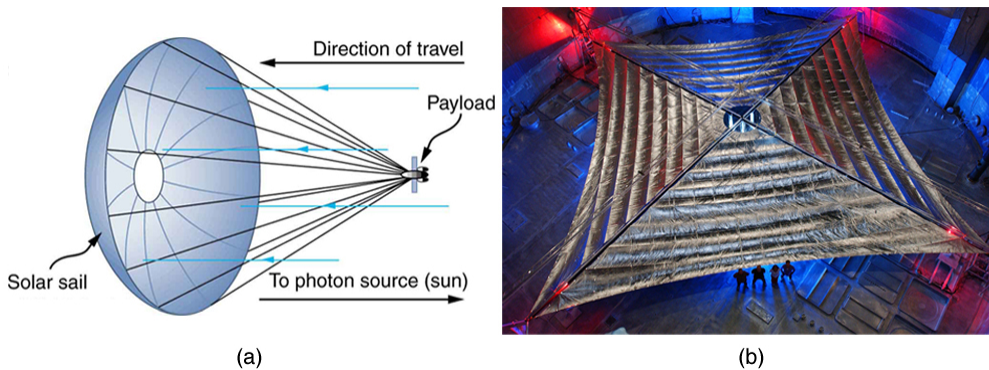
$$E = \frac{hc}{\lambda} = \frac{1240 \text{ eV} \cdot \text{nm}}{500 \text{ nm}} = 2.48 \text{ eV},$$

which is about five orders of magnitude greater.

Discussion

Photon momentum is indeed small. Even if we have huge numbers of them, the total momentum they carry is small. An electron with the same momentum has a 1460 m/s velocity, which is clearly nonrelativistic. A more massive particle with the same momentum would have an even smaller velocity. This is borne out by the fact that it takes far less energy to give an electron the same momentum as a photon. But on a quantum-mechanical scale, especially for high-energy photons interacting with small

masses, photon momentum is significant. Even on a large scale, photon momentum can have an effect if there are enough of them and if there is nothing to prevent the slow recoil of matter. Comet tails are one example, but there are also proposals to build space sails that use huge low-mass mirrors (made of aluminized Mylar) to reflect sunlight. In the vacuum of space, the mirrors would gradually recoil and could actually take spacecraft from place to place in the solar system. (See [\[link\]](#).)



(a) Space sails have been proposed that use the momentum of sunlight reflecting from gigantic low-mass sails to propel spacecraft about the solar system. A Russian test model of this (the Cosmos 1) was launched in 2005, but did not make it into orbit due to a rocket failure. (b) A U.S. version of this, labeled LightSail-1, is scheduled for trial launches in the first part of this decade. It will have a 40-m² sail. (credit: Kim Newton/NASA)

Relativistic Photon Momentum

There is a relationship between photon momentum p and photon energy E that is consistent with the relation given previously for the relativistic total energy of a particle as $E^2 = (pc)^2 + (mc)^2$. We know m is zero for a photon, but p is not, so that $E^2 = (pc)^2 + (mc)^2$ becomes

Equation:

$$E = pc,$$

or

Equation:

$$p = \frac{E}{c} \text{ (photons).}$$

To check the validity of this relation, note that $E = hc/\lambda$ for a photon. Substituting this into $p = E/c$ yields

Equation:

$$p = (hc/\lambda)/c = \frac{h}{\lambda},$$

as determined experimentally and discussed above. Thus, $p = E/c$ is equivalent to Compton's result $p = h/\lambda$. For a further verification of the relationship between photon energy and momentum, see [\[link\]](#).

Note:

Photon Detectors

Almost all detection systems talked about thus far—eyes, photographic plates, photomultiplier tubes in microscopes, and CCD cameras—rely on particle-like properties of photons interacting with a sensitive area. A change is caused and either the change is cascaded or zillions of points are recorded to form an image we detect. These detectors are used in biomedical imaging systems, and there is ongoing research into improving the efficiency of receiving photons, particularly by cooling detection systems and reducing thermal effects.

Example:**Photon Energy and Momentum**

Show that $p = E/c$ for the photon considered in the [\[link\]](#).

Strategy

We will take the energy E found in [\[link\]](#), divide it by the speed of light, and see if the same momentum is obtained as before.

Solution

Given that the energy of the photon is 2.48 eV and converting this to joules, we get

Equation:

$$p = \frac{E}{c} = \frac{(2.48 \text{ eV})(1.60 \times 10^{-19} \text{ J/eV})}{3.00 \times 10^8 \text{ m/s}} = 1.33 \times 10^{-27} \text{ kg} \cdot \text{m/s}.$$

Discussion

This value for momentum is the same as found before (note that unrounded values are used in all calculations to avoid even small rounding errors), an expected verification of the relationship $p = E/c$. This also means the relationship between energy, momentum, and mass given by $E^2 = (pc)^2 + (mc)^2$ applies to both matter and photons. Once again, note that p is not zero, even when m is.

Note:**Problem-Solving Suggestion**

Note that the forms of the constants $h = 4.14 \times 10^{-15} \text{ eV} \cdot \text{s}$ and $hc = 1240 \text{ eV} \cdot \text{nm}$ may be particularly useful for this section's Problems and Exercises.

Section Summary

- Photons have momentum, given by $p = \frac{h}{\lambda}$, where λ is the photon wavelength.

- Photon energy and momentum are related by $p = \frac{E}{c}$, where $E = hf = hc/\lambda$ for a photon.

Conceptual Questions

Exercise:

Problem:

Which formula may be used for the momentum of all particles, with or without mass?

Exercise:

Problem:

Is there any measurable difference between the momentum of a photon and the momentum of matter?

Exercise:

Problem:

Why don't we feel the momentum of sunlight when we are on the beach?

Problems & Exercises

Exercise:

Problem:

- (a) Find the momentum of a 4.00-cm-wavelength microwave photon.
- (b) Discuss why you expect the answer to (a) to be very small.

Solution:

- (a) $1.66 \times 10^{-32} \text{ kg} \cdot \text{m/s}$

(b) The wavelength of microwave photons is large, so the momentum they carry is very small.

Exercise:

Problem:

(a) What is the momentum of a 0.0100-nm-wavelength photon that could detect details of an atom? (b) What is its energy in MeV?

Exercise:

Problem:

(a) What is the wavelength of a photon that has a momentum of $5.00 \times 10^{-29} \text{ kg} \cdot \text{m/s}$? (b) Find its energy in eV.

Solution:

(a) $13.3 \text{ } \mu\text{m}$

(b) $9.38 \times 10^{-2} \text{ eV}$

Exercise:

Problem:

(a) A γ -ray photon has a momentum of $8.00 \times 10^{-21} \text{ kg} \cdot \text{m/s}$. What is its wavelength? (b) Calculate its energy in MeV.

Exercise:

Problem:

(a) Calculate the momentum of a photon having a wavelength of $2.50 \text{ } \mu\text{m}$. (b) Find the velocity of an electron having the same momentum. (c) What is the kinetic energy of the electron, and how does it compare with that of the photon?

Solution:

(a) $2.65 \times 10^{-28} \text{ kg} \cdot \text{m/s}$

(b) 291 m/s

(c) electron 3.86×10^{-26} J, photon 7.96×10^{-20} J, ratio 2.06×10^6

Exercise:

Problem:

Repeat the previous problem for a 10.0-nm-wavelength photon.

Exercise:

Problem:

(a) Calculate the wavelength of a photon that has the same momentum as a proton moving at 1.00% of the speed of light. (b) What is the energy of the photon in MeV? (c) What is the kinetic energy of the proton in MeV?

Solution:

(a) 1.32×10^{-13} m

(b) 9.39 MeV

(c) 4.70×10^{-2} MeV

Exercise:

Problem:

(a) Find the momentum of a 100-keV x-ray photon. (b) Find the equivalent velocity of a neutron with the same momentum. (c) What is the neutron's kinetic energy in keV?

Exercise:

Problem:

Take the ratio of relativistic rest energy, $E = \gamma mc^2$, to relativistic momentum, $p = \gamma mu$, and show that in the limit that mass approaches zero, you find $E/p = c$.

Solution:

$E = \gamma mc^2$ and $P = \gamma mu$, so

Equation:

$$\frac{E}{P} = \frac{\gamma mc^2}{\gamma mu} = \frac{c^2}{u}.$$

As the mass of particle approaches zero, its velocity u will approach c , so that the ratio of energy to momentum in this limit is

Equation:

$$\lim_{m \rightarrow 0} \frac{E}{P} = \frac{c^2}{c} = c$$

which is consistent with the equation for photon energy.

Exercise:**Problem: Construct Your Own Problem**

Consider a space sail such as mentioned in [\[link\]](#). Construct a problem in which you calculate the light pressure on the sail in N/m^2 produced by reflecting sunlight. Also calculate the force that could be produced and how much effect that would have on a spacecraft. Among the things to be considered are the intensity of sunlight, its average wavelength, the number of photons per square meter this implies, the area of the space sail, and the mass of the system being accelerated.

Exercise:**Problem: Unreasonable Results**

A car feels a small force due to the light it sends out from its headlights, equal to the momentum of the light divided by the time in which it is emitted. (a) Calculate the power of each headlight, if they

exert a total force of 2.00×10^{-2} N backward on the car. (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) 3.00×10^6 W

(b) Headlights are way too bright.

(c) Force is too large.

Glossary

photon momentum

the amount of momentum a photon has, calculated by $p = \frac{h}{\lambda} = \frac{E}{c}$

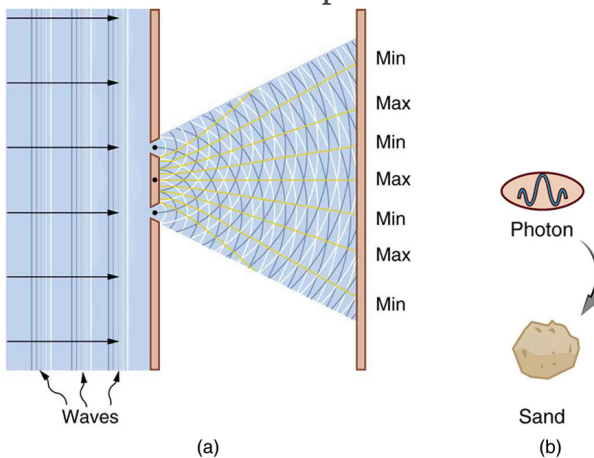
Compton effect

the phenomenon whereby x rays scattered from materials have decreased energy

The Particle-Wave Duality

- Explain what the term particle-wave duality means, and why it is applied to EM radiation.

We have long known that EM radiation is a wave, capable of interference and diffraction. We now see that light can be modeled as photons, which are massless particles. This may seem contradictory, since we ordinarily deal with large objects that never act like both wave and particle. An ocean wave, for example, looks nothing like a rock. To understand small-scale phenomena, we make analogies with the large-scale phenomena we observe directly. When we say something behaves like a wave, we mean it shows interference effects analogous to those seen in overlapping water waves. (See [\[link\]](#).) Two examples of waves are sound and EM radiation. When we say something behaves like a particle, we mean that it interacts as a discrete unit with no interference effects. Examples of particles include electrons, atoms, and photons of EM radiation. How do we talk about a phenomenon that acts like both a particle and a wave?



(a) The interference pattern for light through a double slit is a wave property understood by analogy to water waves. (b) The properties of photons having quantized energy and momentum and acting as a concentrated unit are

understood by analogy to
macroscopic particles.

There is no doubt that EM radiation interferes and has the properties of wavelength and frequency. There is also no doubt that it behaves as particles—photons with discrete energy. We call this twofold nature the **particle-wave duality**, meaning that EM radiation has both particle and wave properties. This so-called duality is simply a term for properties of the photon analogous to phenomena we can observe directly, on a macroscopic scale. If this term seems strange, it is because we do not ordinarily observe details on the quantum level directly, and our observations yield either particle *or* wavelike properties, but never both simultaneously.

Since we have a particle-wave duality for photons, and since we have seen connections between photons and matter in that both have momentum, it is reasonable to ask whether there is a particle-wave duality for matter as well. If the EM radiation we once thought to be a pure wave has particle properties, is it possible that matter has wave properties? The answer is yes. The consequences are tremendous, as we will begin to see in the next section.

Note:

PhET Explorations: Quantum Wave Interference

When do photons, electrons, and atoms behave like particles and when do they behave like waves? Watch waves spread out and interfere as they pass through a double slit, then get detected on a screen as tiny dots. Use quantum detectors to explore how measurements change the waves and the patterns they produce on the screen. [Click to open media in new browser.](#)

Section Summary

- EM radiation can behave like either a particle or a wave.

- This is termed particle-wave duality.

Glossary

particle-wave duality

the property of behaving like either a particle or a wave; the term for the phenomenon that all particles have wave characteristics

The Wave Nature of Matter

- Describe the Davisson-Germer experiment, and explain how it provides evidence for the wave nature of electrons.

De Broglie Wavelength

In 1923 a French physics graduate student named Prince Louis-Victor de Broglie (1892–1987) made a radical proposal based on the hope that nature is symmetric. If EM radiation has both particle and wave properties, then nature would be symmetric if matter also had both particle and wave properties. If what we once thought of as an unequivocal wave (EM radiation) is also a particle, then what we think of as an unequivocal particle (matter) may also be a wave. De Broglie's suggestion, made as part of his doctoral thesis, was so radical that it was greeted with some skepticism. A copy of his thesis was sent to Einstein, who said it was not only probably correct, but that it might be of fundamental importance. With the support of Einstein and a few other prominent physicists, de Broglie was awarded his doctorate.

De Broglie took both relativity and quantum mechanics into account to develop the proposal that *all particles have a wavelength*, given by

Equation:

$$\lambda = \frac{h}{p} \text{ (matter and photons),}$$

where h is Planck's constant and p is momentum. This is defined to be the **de Broglie wavelength**. (Note that we already have this for photons, from the equation $p = h/\lambda$.) The hallmark of a wave is interference. If matter is a wave, then it must exhibit constructive and destructive interference. Why isn't this ordinarily observed? The answer is that in order to see significant interference effects, a wave must interact with an object about the same size as its wavelength. Since h is very small, λ is also small, especially for macroscopic objects. A 3-kg bowling ball moving at 10 m/s, for example, has

Equation:

$$\lambda = h/p = (6.63 \times 10^{-34} \text{ J}\cdot\text{s})/[(3 \text{ kg})(10 \text{ m/s})] = 2 \times 10^{-35} \text{ m}.$$

This means that to see its wave characteristics, the bowling ball would have to interact with something about 10^{-35} m in size—far smaller than anything known. When waves interact with objects much larger than their wavelength, they show negligible interference effects and move in straight lines (such as light rays in geometric optics). To get easily observed interference effects from particles of matter, the longest wavelength and hence smallest mass possible would be useful. Therefore, this effect was first observed with electrons.

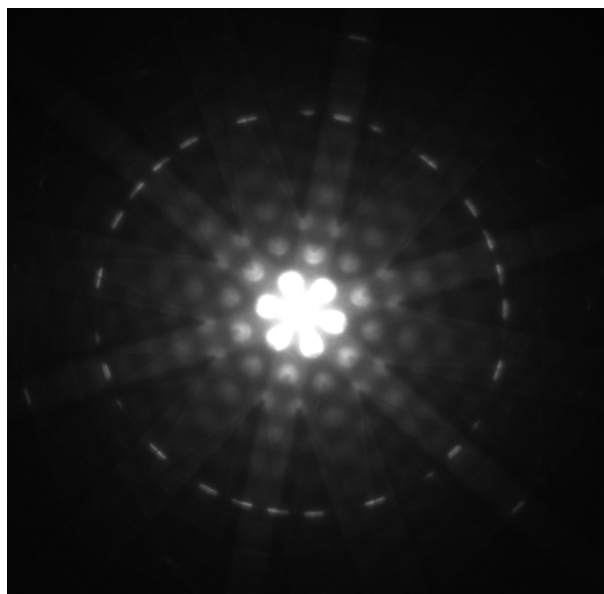
American physicists Clinton J. Davisson and Lester H. Germer in 1925 and, independently, British physicist G. P. Thomson (son of J. J. Thomson, discoverer of the electron) in 1926 scattered electrons from crystals and found diffraction patterns. These patterns are exactly consistent with interference of electrons having the de Broglie wavelength and are somewhat analogous to light interacting with a diffraction grating. (See [\[link\]](#).)

Note:**Connections: Waves**

All microscopic particles, whether massless, like photons, or having mass, like electrons, have wave properties. The relationship between momentum and wavelength is fundamental for all particles.

De Broglie's proposal of a wave nature for all particles initiated a remarkably productive era in which the foundations for quantum mechanics were laid. In 1926, the Austrian physicist Erwin Schrödinger (1887–1961) published four papers in which the wave nature of particles was treated explicitly with wave equations. At the same time, many others began important work. Among them was German physicist Werner Heisenberg

(1901–1976) who, among many other contributions to quantum mechanics, formulated a mathematical treatment of the wave nature of matter that used matrices rather than wave equations. We will deal with some specifics in later sections, but it is worth noting that de Broglie's work was a watershed for the development of quantum mechanics. De Broglie was awarded the Nobel Prize in 1929 for his vision, as were Davisson and G. P. Thomson in 1937 for their experimental verification of de Broglie's hypothesis.



This diffraction pattern was obtained for electrons diffracted by crystalline silicon. Bright regions are those of constructive interference, while dark regions are those of destructive interference. (credit: Ndtthe, Wikimedia Commons)

Example:

Electron Wavelength versus Velocity and Energy

For an electron having a de Broglie wavelength of 0.167 nm (appropriate for interacting with crystal lattice structures that are about this size): (a)

Calculate the electron's velocity, assuming it is nonrelativistic. (b)

Calculate the electron's kinetic energy in eV.

Strategy

For part (a), since the de Broglie wavelength is given, the electron's velocity can be obtained from $\lambda = h/p$ by using the nonrelativistic formula for momentum, $p = mv$. For part (b), once v is obtained (and it has been verified that v is nonrelativistic), the classical kinetic energy is simply $(1/2)mv^2$.

Solution for (a)

Substituting the nonrelativistic formula for momentum ($p = mv$) into the de Broglie wavelength gives

Equation:

$$\lambda = \frac{h}{p} = \frac{h}{mv}.$$

Solving for v gives

Equation:

$$v = \frac{h}{m\lambda}.$$

Substituting known values yields

Equation:

$$v = \frac{6.63 \times 10^{-34} \text{ J} \cdot \text{s}}{(9.11 \times 10^{-31} \text{ kg})(0.167 \times 10^{-9} \text{ m})} = 4.36 \times 10^6 \text{ m/s}.$$

Solution for (b)

While fast compared with a car, this electron's speed is not highly relativistic, and so we can comfortably use the classical formula to find the electron's kinetic energy and convert it to eV as requested.

Equation:

$$\begin{aligned}
 \text{KE} &= \frac{1}{2}mv^2 \\
 &= \frac{1}{2}(9.11 \times 10^{-31} \text{ kg})(4.36 \times 10^6 \text{ m/s})^2 \\
 &= (86.4 \times 10^{-18} \text{ J}) \left(\frac{1 \text{ eV}}{1.602 \times 10^{-19} \text{ J}} \right) \\
 &= 54.0 \text{ eV}
 \end{aligned}$$

Discussion

This low energy means that these 0.167-nm electrons could be obtained by accelerating them through a 54.0-V electrostatic potential, an easy task. The results also confirm the assumption that the electrons are nonrelativistic, since their velocity is just over 1% of the speed of light and the kinetic energy is about 0.01% of the rest energy of an electron (0.511 MeV). If the electrons had turned out to be relativistic, we would have had to use more involved calculations employing relativistic formulas.

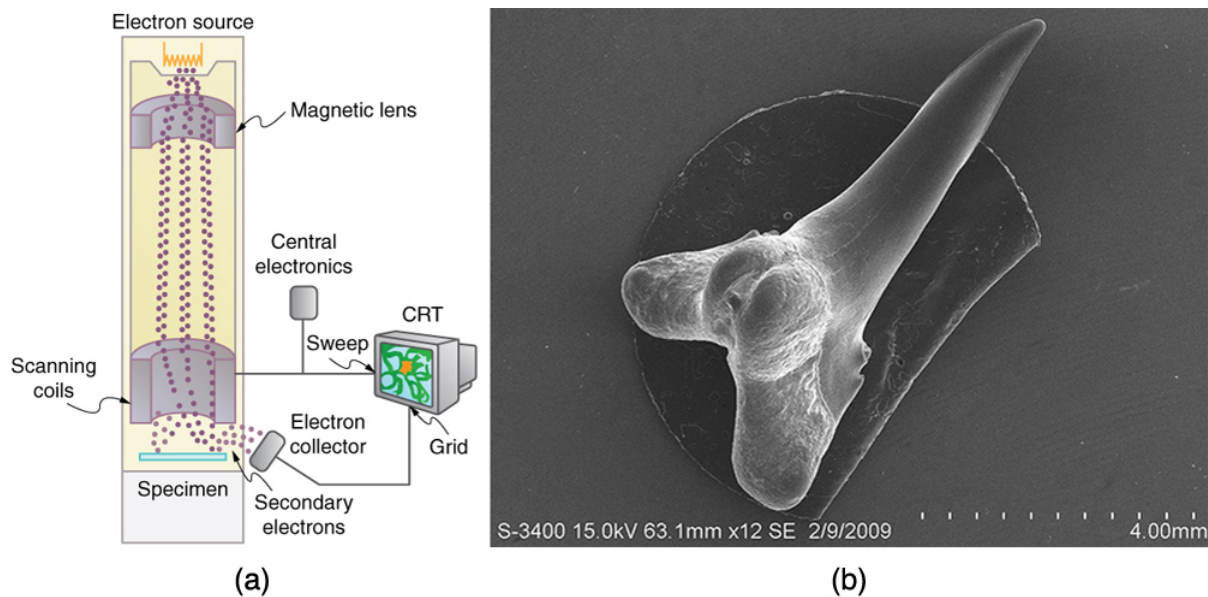
Electron Microscopes

One consequence or use of the wave nature of matter is found in the electron microscope. As we have discussed, there is a limit to the detail observed with any probe having a wavelength. Resolution, or observable detail, is limited to about one wavelength. Since a potential of only 54 V can produce electrons with sub-nanometer wavelengths, it is easy to get electrons with much smaller wavelengths than those of visible light (hundreds of nanometers). Electron microscopes can, thus, be constructed to detect much smaller details than optical microscopes. (See [\[link\]](#).)

There are basically two types of electron microscopes. The transmission electron microscope (TEM) accelerates electrons that are emitted from a hot filament (the cathode). The beam is broadened and then passes through the sample. A magnetic lens focuses the beam image onto a fluorescent screen, a photographic plate, or (most probably) a CCD (light sensitive camera), from which it is transferred to a computer. The TEM is similar to the optical microscope, but it requires a thin sample examined in a vacuum. However it can resolve details as small as 0.1 nm (10^{-10} m), providing magnifications

of 100 million times the size of the original object. The TEM has allowed us to see individual atoms and structure of cell nuclei.

The scanning electron microscope (SEM) provides images by using secondary electrons produced by the primary beam interacting with the surface of the sample (see [\[link\]](#)). The SEM also uses magnetic lenses to focus the beam onto the sample. However, it moves the beam around electrically to “scan” the sample in the x and y directions. A CCD detector is used to process the data for each electron position, producing images like the one at the beginning of this chapter. The SEM has the advantage of not requiring a thin sample and of providing a 3-D view. However, its resolution is about ten times less than a TEM.



Schematic of a scanning electron microscope (SEM) (a) used to observe small details, such as those seen in this image of a tooth of a *Himipristis*, a type of shark (b). (credit: Dallas Krentzel, Flickr)

Electrons were the first particles with mass to be directly confirmed to have the wavelength proposed by de Broglie. Subsequently, protons, helium nuclei, neutrons, and many others have been observed to exhibit

interference when they interact with objects having sizes similar to their de Broglie wavelength. The de Broglie wavelength for massless particles was well established in the 1920s for photons, and it has since been observed that all massless particles have a de Broglie wavelength $\lambda = h/p$. The wave nature of all particles is a universal characteristic of nature. We shall see in following sections that implications of the de Broglie wavelength include the quantization of energy in atoms and molecules, and an alteration of our basic view of nature on the microscopic scale. The next section, for example, shows that there are limits to the precision with which we may make predictions, regardless of how hard we try. There are even limits to the precision with which we may measure an object's location or energy.

Note:

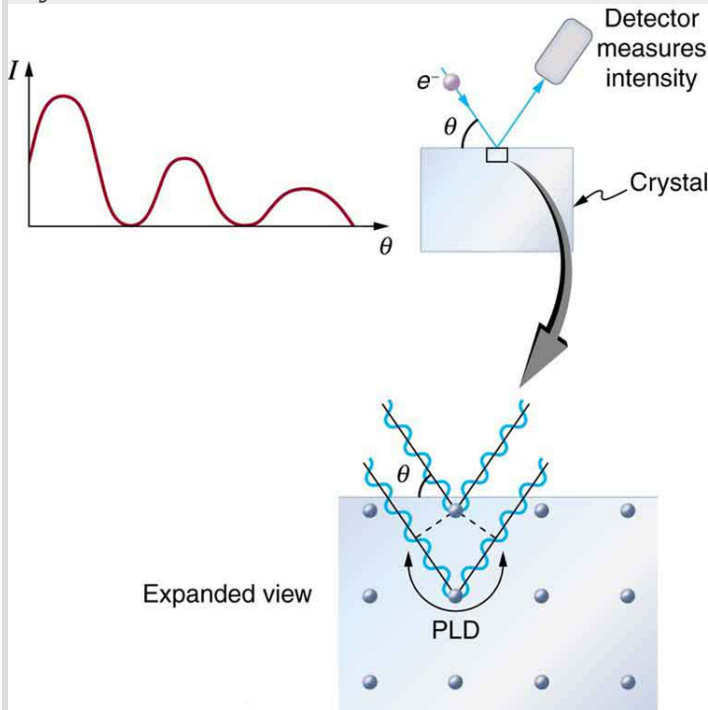
Making Connections: A Submicroscopic Diffraction Grating

The wave nature of matter allows it to exhibit all the characteristics of other, more familiar, waves. Diffraction gratings, for example, produce diffraction patterns for light that depend on grating spacing and the wavelength of the light. This effect, as with most wave phenomena, is most pronounced when the wave interacts with objects having a size similar to its wavelength. For gratings, this is the spacing between multiple slits.)

When electrons interact with a system having a spacing similar to the electron wavelength, they show the same types of interference patterns as light does for diffraction gratings, as shown at top left in [\[link\]](#).

Atoms are spaced at regular intervals in a crystal as parallel planes, as shown in the bottom part of [\[link\]](#). The spacings between these planes act like the openings in a diffraction grating. At certain incident angles, the paths of electrons scattering from successive planes differ by one wavelength and, thus, interfere constructively. At other angles, the path length differences are not an integral wavelength, and there is partial to total destructive interference. This type of scattering from a large crystal with well-defined lattice planes can produce dramatic interference patterns. It is called *Bragg reflection*, for the father-and-son team who first explored and analyzed it in some detail. The expanded view also shows the path-length differences and indicates how these depend on incident angle θ in a

manner similar to the diffraction patterns for x rays reflecting from a crystal.



The diffraction pattern at top left is produced by scattering electrons from a crystal and is graphed as a function of incident angle relative to the regular array of atoms in a crystal, as shown at bottom. Electrons scattering from the second layer of atoms travel farther than those scattered from the top layer. If the path length difference (PLD) is an integral wavelength, there is constructive interference.

Let us take the spacing between parallel planes of atoms in the crystal to be d . As mentioned, if the path length difference (PLD) for the electrons is a whole number of wavelengths, there will be constructive interference—that is, $\text{PLD} = n\lambda$ ($n = 1, 2, 3, \dots$). Because $AB = BC = d \sin \theta$, we have constructive interference when $n\lambda = 2d \sin \theta$. This relationship is

called the *Bragg equation* and applies not only to electrons but also to x rays.

The wavelength of matter is a submicroscopic characteristic that explains a macroscopic phenomenon such as Bragg reflection. Similarly, the wavelength of light is a submicroscopic characteristic that explains the macroscopic phenomenon of diffraction patterns.

Section Summary

- Particles of matter also have a wavelength, called the de Broglie wavelength, given by $\lambda = \frac{h}{p}$, where p is momentum.
- Matter is found to have the same *interference characteristics* as any other wave.

Conceptual Questions

Exercise:

Problem:

How does the interference of water waves differ from the interference of electrons? How are they analogous?

Exercise:

Problem: Describe one type of evidence for the wave nature of matter.

Exercise:

Problem:

Describe one type of evidence for the particle nature of EM radiation.

Problems & Exercises

Exercise:

Problem:

At what velocity will an electron have a wavelength of 1.00 m?

Solution:

$$7.28 \times 10^{-4} \text{ m}$$

Exercise:**Problem:**

What is the wavelength of an electron moving at 3.00% of the speed of light?

Exercise:**Problem:**

At what velocity does a proton have a 6.00-fm wavelength (about the size of a nucleus)? Assume the proton is nonrelativistic. (1 femtometer = 10^{-15} m.)

Solution:

$$6.62 \times 10^7 \text{ m/s}$$

Exercise:**Problem:**

What is the velocity of a 0.400-kg billiard ball if its wavelength is 7.50 cm (large enough for it to interfere with other billiard balls)?

Exercise:**Problem:**

Find the wavelength of a proton moving at 1.00% of the speed of light.

Solution:

$$1.32 \times 10^{-13} \text{ m}$$

Exercise:**Problem:**

Experiments are performed with ultracold neutrons having velocities as small as 1.00 m/s. (a) What is the wavelength of such a neutron? (b) What is its kinetic energy in eV?

Exercise:**Problem:**

(a) Find the velocity of a neutron that has a 6.00-fm wavelength (about the size of a nucleus). Assume the neutron is nonrelativistic. (b) What is the neutron's kinetic energy in MeV?

Solution:

(a) $6.62 \times 10^7 \text{ m/s}$

(b) 22.9 MeV

Exercise:**Problem:**

What is the wavelength of an electron accelerated through a 30.0-kV potential, as in a TV tube?

Exercise:**Problem:**

What is the kinetic energy of an electron in a TEM having a 0.0100-nm wavelength?

Solution:

Equation: 15.1 keV

Exercise:

Problem:

(a) Calculate the velocity of an electron that has a wavelength of $1.00\text{ }\mu\text{m}$. (b) Through what voltage must the electron be accelerated to have this velocity?

Exercise:**Problem:**

The velocity of a proton emerging from a Van de Graaff accelerator is 25.0% of the speed of light. (a) What is the proton's wavelength? (b) What is its kinetic energy, assuming it is nonrelativistic? (c) What was the equivalent voltage through which it was accelerated?

Solution:

(a) 5.29 fm

(b) $4.70 \times 10^{-12}\text{ J}$

(c) 29.4 MV

Exercise:**Problem:**

The kinetic energy of an electron accelerated in an x-ray tube is 100 keV . Assuming it is nonrelativistic, what is its wavelength?

Exercise:**Problem: Unreasonable Results**

(a) Assuming it is nonrelativistic, calculate the velocity of an electron with a 0.100-fm wavelength (small enough to detect details of a nucleus). (b) What is unreasonable about this result? (c) Which assumptions are unreasonable or inconsistent?

Solution:

(a) $7.28 \times 10^{12} \text{ m/s}$

(b) This is thousands of times the speed of light (an impossibility).

(c) The assumption that the electron is non-relativistic is unreasonable at this wavelength.

Glossary

de Broglie wavelength

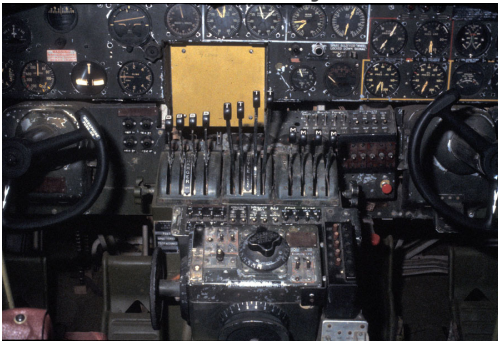
the wavelength possessed by a particle of matter, calculated by

$$\lambda = h/p$$

Nuclear Radioactivity

- Explain nuclear radiation.
- Explain the types of radiation—alpha emission, beta emission, and gamma emission.
- Explain the ionization of radiation in an atom.
- Define the range of radiation.

The discovery and study of nuclear radioactivity quickly revealed evidence of revolutionary new physics. In addition, uses for nuclear radiation also emerged quickly—for example, people such as Ernest Rutherford used it to determine the size of the nucleus and devices were painted with radon-doped paint to make them glow in the dark (see [\[link\]](#)). We therefore begin our study of nuclear physics with the discovery and basic features of nuclear radioactivity.



The dials of this World War II aircraft glow in the dark, because they are painted with radium-doped phosphorescent paint. It is a poignant reminder of the dual nature of radiation. Although radium paint dials are conveniently visible day and night, they emit radon, a radioactive gas that is hazardous and is not

directly sensed. (credit:
U.S. Air Force Photo)

Discovery of Nuclear Radioactivity

In 1896, the French physicist Antoine Henri Becquerel (1852–1908) accidentally found that a uranium-rich mineral called pitchblende emits invisible, penetrating rays that can darken a photographic plate enclosed in an opaque envelope. The rays therefore carry energy; but amazingly, the pitchblende emits them continuously without any energy input. This is an apparent violation of the law of conservation of energy, one that we now understand is due to the conversion of a small amount of mass into energy, as related in Einstein's famous equation $E = mc^2$. It was soon evident that Becquerel's rays originate in the nuclei of the atoms and have other unique characteristics. The emission of these rays is called **nuclear radioactivity** or simply **radioactivity**. The rays themselves are called **nuclear radiation**. A nucleus that spontaneously destroys part of its mass to emit radiation is said to **decay** (a term also used to describe the emission of radiation by atoms in excited states). A substance or object that emits nuclear radiation is said to be **radioactive**.

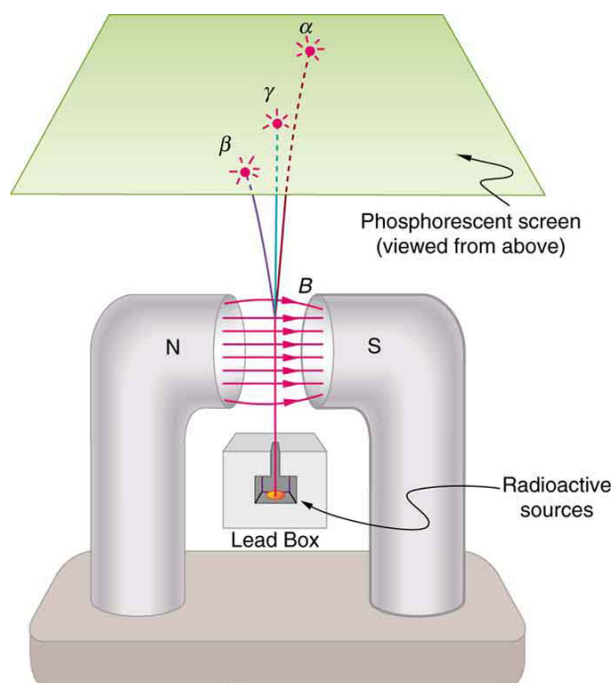
Two types of experimental evidence imply that Becquerel's rays originate deep in the heart (or nucleus) of an atom. First, the radiation is found to be associated with certain elements, such as uranium. Radiation does not vary with chemical state—that is, uranium is radioactive whether it is in the form of an element or compound. In addition, radiation does not vary with temperature, pressure, or ionization state of the uranium atom. Since all of these factors affect electrons in an atom, the radiation cannot come from electron transitions, as atomic spectra do. The huge energy emitted during each event is the second piece of evidence that the radiation cannot be atomic. Nuclear radiation has energies of the order of 10^6 eV per event, which is much greater than the typical atomic energies (a few eV), such as that observed in spectra and chemical reactions, and more than ten times as high as the most energetic characteristic x rays. Becquerel did not vigorously pursue his discovery for very long. In 1898, Marie Curie (1867–

1934), then a graduate student married to the already well-known French physicist Pierre Curie (1859–1906), began her doctoral study of Becquerel's rays. She and her husband soon discovered two new radioactive elements, which she named *polonium* (after her native land) and *radium* (because it radiates). These two new elements filled holes in the periodic table and, further, displayed much higher levels of radioactivity per gram of material than uranium. Over a period of four years, working under poor conditions and spending their own funds, the Curies processed more than a ton of uranium ore to isolate a gram of radium salt. Radium became highly sought after, because it was about two million times as radioactive as uranium. Curie's radium salt glowed visibly from the radiation that took its toll on them and other unaware researchers. Shortly after completing her Ph.D., both Curies and Becquerel shared the 1903 Nobel Prize in physics for their work on radioactivity. Pierre was killed in a horse cart accident in 1906, but Marie continued her study of radioactivity for nearly 30 more years. Awarded the 1911 Nobel Prize in chemistry for her discovery of two new elements, she remains the only person to win Nobel Prizes in physics and chemistry. Marie's radioactive fingerprints on some pages of her notebooks can still expose film, and she suffered from radiation-induced lesions. She died of leukemia likely caused by radiation, but she was active in research almost until her death in 1934. The following year, her daughter and son-in-law, Irene and Frederic Joliot-Curie, were awarded the Nobel Prize in chemistry for their discovery of artificially induced radiation, adding to a remarkable family legacy.

Alpha, Beta, and Gamma

Research begun by people such as New Zealander Ernest Rutherford soon after the discovery of nuclear radiation indicated that different types of rays are emitted. Eventually, three types were distinguished and named **alpha** (α), **beta** (β), and **gamma** (γ), because, like x-rays, their identities were initially unknown. [\[link\]](#) shows what happens if the rays are passed through a magnetic field. The γ s are unaffected, while the α s and β s are deflected in opposite directions, indicating the α s are positive, the β s negative, and the γ s uncharged. Rutherford used both magnetic and electric fields to show that α s have a positive charge twice the magnitude of an electron, or $+2 |q_e|$. In the process, he found the α s charge to mass ratio to be several

thousand times smaller than the electron's. Later on, Rutherford collected α s from a radioactive source and passed an electric discharge through them, obtaining the spectrum of recently discovered helium gas. Among many important discoveries made by Rutherford and his collaborators was the proof that *α radiation is the emission of a helium nucleus*. Rutherford won the Nobel Prize in chemistry in 1908 for his early work. He continued to make important contributions until his death in 1934.



Alpha, beta, and gamma rays are passed through a magnetic field on the way to a phosphorescent screen. The α s and β s bend in opposite directions, while the γ s are unaffected, indicating a positive charge for α s, negative for β s, and neutral for γ s. Consistent results are obtained with electric fields. Collection of the radiation offers further

confirmation from the direct measurement of excess charge.

Other researchers had already proved that β s are negative and have the same mass and same charge-to-mass ratio as the recently discovered electron. By 1902, it was recognized that *β radiation is the emission of an electron*. Although β s are electrons, they do not exist in the nucleus before it decays and are not ejected atomic electrons—the electron is created in the nucleus at the instant of decay.

Since γ s remain unaffected by electric and magnetic fields, it is natural to think they might be photons. Evidence for this grew, but it was not until 1914 that this was proved by Rutherford and collaborators. By scattering γ radiation from a crystal and observing interference, they demonstrated that *γ radiation is the emission of a high-energy photon by a nucleus*. In fact, γ radiation comes from the de-excitation of a nucleus, just as an x ray comes from the de-excitation of an atom. The names " γ ray" and "x ray" identify the source of the radiation. At the same energy, γ rays and x rays are otherwise identical.

Type of Radiation	Range
α -Particles	A sheet of paper, a few cm of air, fractions of a mm of tissue

Type of Radiation	Range
β -Particles	A thin aluminum plate, or tens of cm of tissue
γ Rays	Several cm of lead or meters of concrete

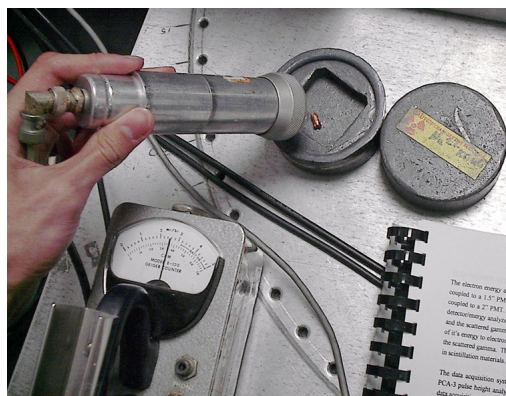
Properties of Nuclear Radiation

Ionization and Range

Two of the most important characteristics of α , β , and γ rays were recognized very early. All three types of nuclear radiation produce *ionization* in materials, but they penetrate different distances in materials—that is, they have different *ranges*. Let us examine why they have these characteristics and what are some of the consequences.

Like x rays, nuclear radiation in the form of α s, β s, and γ s has enough energy per event to ionize atoms and molecules in any material. The energy emitted in various nuclear decays ranges from a few keV to more than 10 MeV, while only a few eV are needed to produce ionization. The effects of x rays and nuclear radiation on biological tissues and other materials, such as solid state electronics, are directly related to the ionization they produce. All of them, for example, can damage electronics or kill cancer cells. In addition, methods for detecting x rays and nuclear radiation are based on ionization, directly or indirectly. All of them can ionize the air between the plates of a capacitor, for example, causing it to discharge. This is the basis of inexpensive personal radiation monitors, such as pictured in [\[link\]](#). Apart from α , β , and γ , there are other forms of nuclear radiation as well, and these also produce ionization with similar effects. We define **ionizing radiation** as any form of radiation that produces ionization

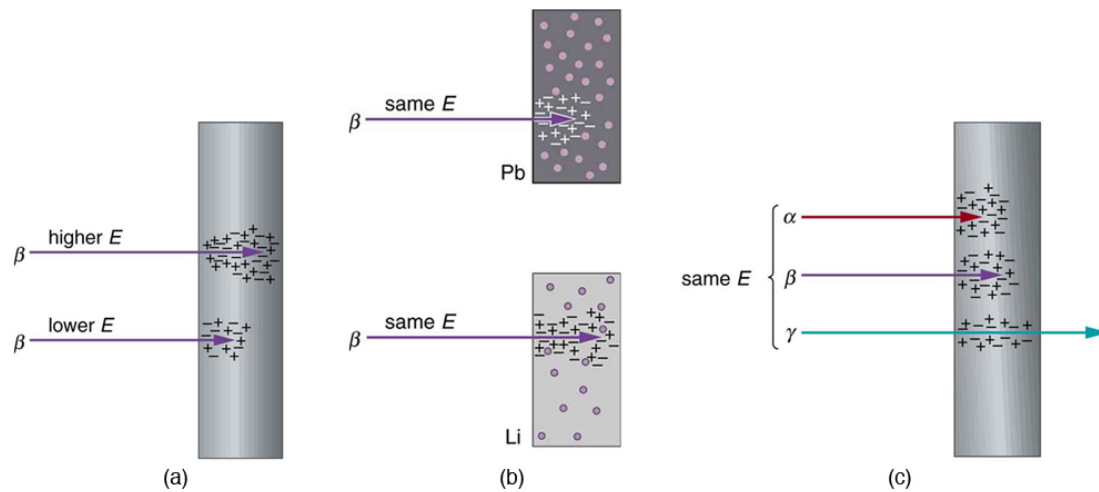
whether nuclear in origin or not, since the effects and detection of the radiation are related to ionization.



These dosimeters (literally, dose meters) are personal radiation monitors that detect the amount of radiation by the discharge of a rechargeable internal capacitor. The amount of discharge is related to the amount of ionizing radiation encountered, a measurement of dose. One dosimeter is shown in the charger. Its scale is read through an eyepiece on the top. (credit: L. Chang, Wikimedia Commons)

The **range of radiation** is defined to be the distance it can travel through a material. Range is related to several factors, including the energy of the

radiation, the material encountered, and the type of radiation (see [\[link\]](#)). The higher the *energy*, the greater the range, all other factors being the same. This makes good sense, since radiation loses its energy in materials primarily by producing ionization in them, and each ionization of an atom or a molecule requires energy that is removed from the radiation. The amount of ionization is, thus, directly proportional to the energy of the particle of radiation, as is its range.



The penetration or range of radiation depends on its energy, the material it encounters, and the type of radiation. (a) Greater energy means greater range. (b) Radiation has a smaller range in materials with high electron density. (c) Alphas have the smallest range, betas have a greater range, and gammas penetrate the farthest.

Radiation can be absorbed or shielded by materials, such as the lead aprons dentists drape on us when taking x rays. Lead is a particularly effective shield compared with other materials, such as plastic or air. How does the range of radiation depend on *material*? Ionizing radiation interacts best with charged particles in a material. Since electrons have small masses, they most readily absorb the energy of the radiation in collisions. The greater the

density of a material and, in particular, the greater the density of electrons within a material, the smaller the range of radiation.

Note:

Collisions

Conservation of energy and momentum often results in energy transfer to a less massive object in a collision. This was discussed in detail in [Work, Energy, and Energy Resources](#), for example.

Different *types* of radiation have different ranges when compared at the same energy and in the same material. Alphas have the shortest range, betas penetrate farther, and gammas have the greatest range. This is directly related to charge and speed of the particle or type of radiation. At a given energy, each α , β , or γ will produce the same number of ionizations in a material (each ionization requires a certain amount of energy on average). The more readily the particle produces ionization, the more quickly it will lose its energy. The effect of *charge* is as follows: The α has a charge of $+2q_e$, the β has a charge of $-q_e$, and the γ is uncharged. The electromagnetic force exerted by the α is thus twice as strong as that exerted by the β and it is more likely to produce ionization. Although chargeless, the γ does interact weakly because it is an electromagnetic wave, but it is less likely to produce ionization in any encounter. More quantitatively, the change in momentum Δp given to a particle in the material is $\Delta p = F\Delta t$, where F is the force the α , β , or γ exerts over a time Δt . The smaller the charge, the smaller is F and the smaller is the momentum (and energy) lost. Since the speed of alphas is about 5% to 10% of the speed of light, classical (non-relativistic) formulas apply.

The *speed* at which they travel is the other major factor affecting the range of α s, β s, and γ s. The faster they move, the less time they spend in the vicinity of an atom or a molecule, and the less likely they are to interact. Since α s and β s are particles with mass (helium nuclei and electrons, respectively), their energy is kinetic, given classically by $\frac{1}{2}mv^2$. The mass

of the β particle is thousands of times less than that of the α s, so that β s must travel much faster than α s to have the same energy. Since β s move faster (most at relativistic speeds), they have less time to interact than α s. Gamma rays are photons, which must travel at the speed of light. They are even less likely to interact than a β , since they spend even less time near a given atom (and they have no charge). The range of γ s is thus greater than the range of β s.

Alpha radiation from radioactive sources has a range much less than a millimeter of biological tissues, usually not enough to even penetrate the dead layers of our skin. On the other hand, the same α radiation can penetrate a few centimeters of air, so mere distance from a source prevents α radiation from reaching us. This makes α radiation relatively safe for our body compared to β and γ radiation. Typical β radiation can penetrate a few millimeters of tissue or about a meter of air. Beta radiation is thus hazardous even when not ingested. The range of β s in lead is about a millimeter, and so it is easy to store β sources in lead radiation-proof containers. Gamma rays have a much greater range than either α s or β s. In fact, if a given thickness of material, like a lead brick, absorbs 90% of the γ s, then a second lead brick will only absorb 90% of what got through the first. Thus, γ s do not have a well-defined range; we can only cut down the amount that gets through. Typically, γ s can penetrate many meters of air, go right through our bodies, and are effectively shielded (that is, reduced in intensity to acceptable levels) by many centimeters of lead. One benefit of γ s is that they can be used as radioactive tracers (see [\[link\]](#)).



This image of the concentration of a radioactive tracer in a patient's body reveals where the most active bone cells are, an indication of bone cancer. A short-lived radioactive substance that locates itself selectively is given to the patient, and the radiation is measured with an external detector. The emitted γ radiation has a sufficient range to leave the body—the range of α s and β s is too small for them to be observed outside the patient. (credit: Kieran Maher, Wikimedia Commons)

Note:**PhET Explorations: Beta Decay**

Build an atom out of protons, neutrons, and electrons, and see how the element, charge, and mass change. Then play a game to test your ideas!

<https://archive.cnx.org/specials/f0a27b96-f5c8-11e5-a22c-73f8c149bebf/beta-decay/#sim-multiple-atoms>

Section Summary

- Some nuclei are radioactive—they spontaneously decay destroying some part of their mass and emitting energetic rays, a process called nuclear radioactivity.
- Nuclear radiation, like x rays, is ionizing radiation, because energy sufficient to ionize matter is emitted in each decay.
- The range (or distance traveled in a material) of ionizing radiation is directly related to the charge of the emitted particle and its energy, with greater-charge and lower-energy particles having the shortest ranges.
- Radiation detectors are based directly or indirectly upon the ionization created by radiation, as are the effects of radiation on living and inert materials.

Conceptual Questions

Exercise:**Problem:**

Suppose the range for 5.0 MeV α ray is known to be 2.0 mm in a certain material. Does this mean that every 5.0 MeV α ray that strikes this material travels 2.0 mm, or does the range have an average value with some statistical fluctuations in the distances traveled? Explain.

Exercise:

Problem:

What is the difference between γ rays and characteristic x rays? Is either necessarily more energetic than the other? Which can be the most energetic?

Exercise:**Problem:**

Ionizing radiation interacts with matter by scattering from electrons and nuclei in the substance. Based on the law of conservation of momentum and energy, explain why electrons tend to absorb more energy than nuclei in these interactions.

Exercise:**Problem:**

What characteristics of radioactivity show it to be nuclear in origin and not atomic?

Exercise:**Problem:**

What is the source of the energy emitted in radioactive decay? Identify an earlier conservation law, and describe how it was modified to take such processes into account.

Exercise:**Problem:**

Consider [\[link\]](#). If an electric field is substituted for the magnetic field with positive charge instead of the north pole and negative charge instead of the south pole, in which directions will the α , β , and γ rays bend?

Exercise:

Problem:

Explain how an α particle can have a larger range in air than a β particle with the same energy in lead.

Exercise:**Problem:**

Arrange the following according to their ability to act as radiation shields, with the best first and worst last. Explain your ordering in terms of how radiation loses its energy in matter.

- (a) A solid material with low density composed of low-mass atoms.
- (b) A gas composed of high-mass atoms.
- (c) A gas composed of low-mass atoms.
- (d) A solid with high density composed of high-mass atoms.

Exercise:**Problem:**

Often, when people have to work around radioactive materials spills, we see them wearing white coveralls (usually a plastic material). What types of radiation (if any) do you think these suits protect the worker from, and how?

Glossary

alpha rays

one of the types of rays emitted from the nucleus of an atom

beta rays

one of the types of rays emitted from the nucleus of an atom

gamma rays

one of the types of rays emitted from the nucleus of an atom

ionizing radiation

radiation (whether nuclear in origin or not) that produces ionization
whether nuclear in origin or not

nuclear radiation

rays that originate in the nuclei of atoms, the first examples of which
were discovered by Becquerel

radioactivity

the emission of rays from the nuclei of atoms

radioactive

a substance or object that emits nuclear radiation

range of radiation

the distance that the radiation can travel through a material

Binding Energy

- Define and discuss binding energy.
- Calculate the binding energy per nucleon of a particle.

The more tightly bound a system is, the stronger the forces that hold it together and the greater the energy required to pull it apart. We can therefore learn about nuclear forces by examining how tightly bound the nuclei are. We define the **binding energy** (BE) of a nucleus to be *the energy required to completely disassemble it into separate protons and neutrons*. We can determine the BE of a nucleus from its rest mass. The two are connected through Einstein's famous relationship $E = (\Delta m)c^2$. A bound system has a *smaller* mass than its separate constituents; the more tightly the nucleons are bound together, the smaller the mass of the nucleus.

Imagine pulling a nuclide apart as illustrated in [\[link\]](#). Work done to overcome the nuclear forces holding the nucleus together puts energy into the system. By definition, the energy input equals the binding energy BE. The pieces are at rest when separated, and so the energy put into them increases their total rest mass compared with what it was when they were glued together as a nucleus. That mass increase is thus $\Delta m = \text{BE}/c^2$. This difference in mass is known as *mass defect*. It implies that the mass of the nucleus is less than the sum of the masses of its constituent protons and neutrons. A nuclide ${}^A\text{X}$ has Z protons and N neutrons, so that the difference in mass is

Equation:

$$\Delta m = (Zm_p + Nm_n) - m_{\text{tot}}.$$

Thus,

Equation:

$$\text{BE} = (\Delta m)c^2 = [(Zm_p + Nm_n) - m_{\text{tot}}]c^2,$$

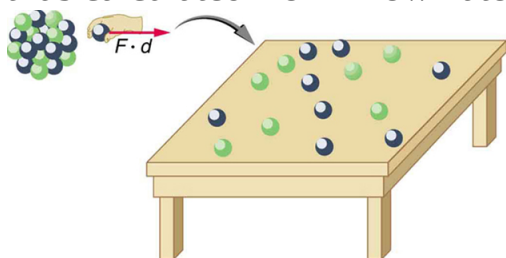
where m_{tot} is the mass of the nuclide ${}^A\text{X}$, m_p is the mass of a proton, and m_n is the mass of a neutron. Traditionally, we deal with the masses of

neutral atoms. To get atomic masses into the last equation, we first add Z electrons to m_{tot} , which gives $m(^A\text{X})$, the atomic mass of the nuclide. We then add Z electrons to the Z protons, which gives $Zm(^1\text{H})$, or Z times the mass of a hydrogen atom. Thus the binding energy of a nuclide ^AX is

Equation:

$$\text{BE} = \left\{ [Zm(^1\text{H}) + Nm_n] - m(^A\text{X}) \right\} c^2.$$

The atomic masses can be found in [Appendix A](#), most conveniently expressed in unified atomic mass units u ($1 u = 931.5 \text{ MeV}/c^2$). BE is thus calculated from known atomic masses.



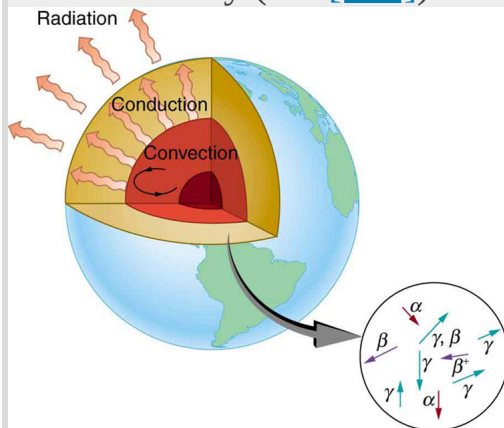
Work done to pull a nucleus apart into its constituent protons and neutrons increases the mass of the system. The work to disassemble the nucleus equals its binding energy BE. A bound system has less mass than the sum of its parts, especially noticeable in the nuclei, where forces and energies are very large.

Note:**Things Great and Small****Nuclear Decay Helps Explain Earth's Hot Interior**

A puzzle created by radioactive dating of rocks is resolved by radioactive heating of Earth's interior. This intriguing story is another example of how small-scale physics can explain large-scale phenomena.

Radioactive dating plays a role in determining the approximate age of the Earth. The oldest rocks on Earth solidified about 3.5×10^9 years ago—a number determined by uranium-238 dating. These rocks could only have solidified once the surface of the Earth had cooled sufficiently. The temperature of the Earth at formation can be estimated based on gravitational potential energy of the assemblage of pieces being converted to thermal energy. Using heat transfer concepts discussed in

[Thermodynamics](#) it is then possible to calculate how long it would take for the surface to cool to rock-formation temperatures. The result is about 10^9 years. The first rocks formed have been solid for 3.5×10^9 years, so that the age of the Earth is approximately 4.5×10^9 years. There is a large body of other types of evidence (both Earth-bound and solar system characteristics are used) that supports this age. The puzzle is that, given its age and initial temperature, the center of the Earth should be much cooler than it is today (see [\[link\]](#)).



The center of the Earth
cools by well-known heat
transfer methods.

Convection in the liquid
regions and conduction

move thermal energy to the surface, where it radiates into cold, dark space. Given the age of the Earth and its initial temperature, it should have cooled to a lower temperature by now. The blowup shows that nuclear decay releases energy in the Earth's interior. This energy has slowed the cooling process and is responsible for the interior still being molten.

We know from seismic waves produced by earthquakes that parts of the interior of the Earth are liquid. Shear or transverse waves cannot travel through a liquid and are not transmitted through the Earth's core. Yet compression or longitudinal waves can pass through a liquid and do go through the core. From this information, the temperature of the interior can be estimated. As noticed, the interior should have cooled more from its initial temperature in the 4.5×10^9 years since its formation. In fact, it should have taken no more than about 10^9 years to cool to its present temperature. What is keeping it hot? The answer seems to be radioactive decay of primordial elements that were part of the material that formed the Earth (see the blowup in [\[link\]](#)).

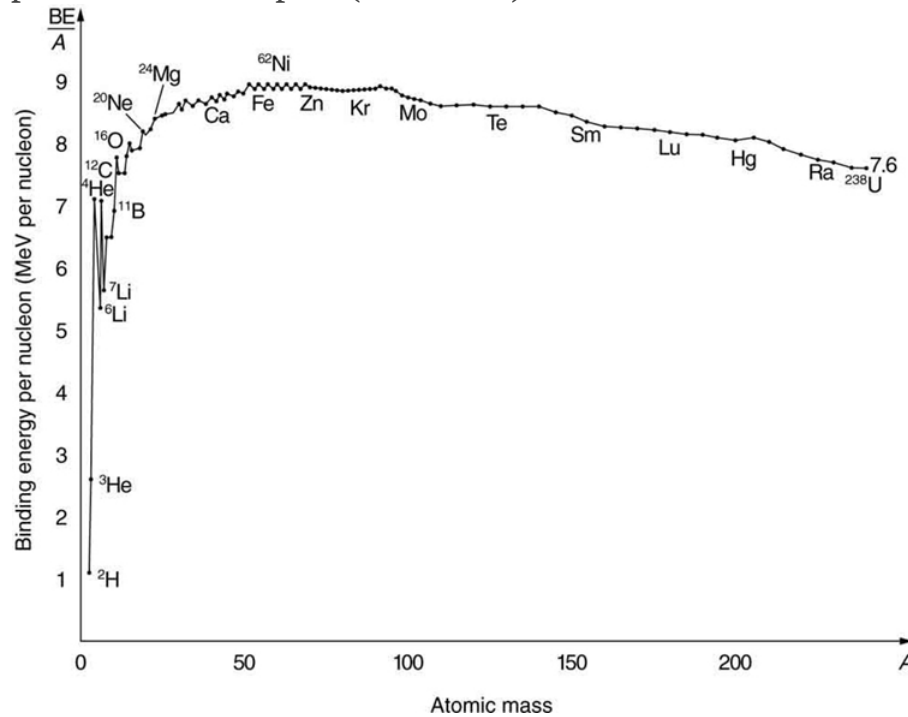
Nuclides such as ^{238}U and ^{40}K have half-lives similar to or longer than the age of the Earth, and their decay still contributes energy to the interior. Some of the primordial radioactive nuclides have unstable decay products that also release energy— ^{238}U has a long decay chain of these. Further, there were more of these primordial radioactive nuclides early in the life of the Earth, and thus the activity and energy contributed were greater then (perhaps by an order of magnitude). The amount of power created by these decays per cubic meter is very small. However, since a huge volume of

material lies deep below the surface, this relatively small amount of energy cannot escape quickly. The power produced near the surface has much less distance to go to escape and has a negligible effect on surface temperatures.

A final effect of this trapped radiation merits mention. Alpha decay produces helium nuclei, which form helium atoms when they are stopped and capture electrons. Most of the helium on Earth is obtained from wells and is produced in this manner. Any helium in the atmosphere will escape in geologically short times because of its high thermal velocity.

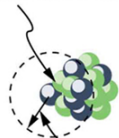
What patterns and insights are gained from an examination of the binding energy of various nuclides? First, we find that BE is approximately proportional to the number of nucleons A in any nucleus. About twice as much energy is needed to pull apart a nucleus like ^{24}Mg compared with pulling apart ^{12}C , for example. To help us look at other effects, we divide BE by A and consider the **binding energy per nucleon**, BE/A . The graph of BE/A in [\[link\]](#) reveals some very interesting aspects of nuclei. We see that the binding energy per nucleon averages about 8 MeV, but is lower for both the lightest and heaviest nuclei. This overall trend, in which nuclei with A equal to about 60 have the greatest BE/A and are thus the most tightly bound, is due to the combined characteristics of the attractive nuclear forces and the repulsive Coulomb force. It is especially important to note two things—the strong nuclear force is about 100 times stronger than the Coulomb force, *and* the nuclear forces are shorter in range compared to the Coulomb force. So, for low-mass nuclei, the nuclear attraction dominates and each added nucleon forms bonds with all others, causing progressively heavier nuclei to have progressively greater values of BE/A . This continues up to $A \approx 60$, roughly corresponding to the mass number of iron. Beyond that, new nucleons added to a nucleus will be too far from some others to feel their nuclear attraction. Added protons, however, feel the repulsion of all other protons, since the Coulomb force is longer in range. Coulomb repulsion grows for progressively heavier nuclei, but nuclear attraction remains about the same, and so BE/A becomes smaller. This is why stable nuclei heavier than $A \approx 40$ have more neutrons than

protons. Coulomb repulsion is reduced by having more neutrons to keep the protons farther apart (see [\[link\]](#)).



A graph of average binding energy per nucleon, BE/A , for stable nuclei. The most tightly bound nuclei are those with A near 60, where the attractive nuclear force has its greatest effect. At higher A s, the Coulomb repulsion progressively reduces the binding energy per nucleon, because the nuclear force is short ranged. The spikes on the curve are very tightly bound nuclides and indicate shell closures.

Nucleons inside range
feel nuclear force directly



Range of nuclear force

The nuclear force is

attractive and stronger than the Coulomb force, but it is short ranged. In low-mass nuclei, each nucleon feels the nuclear attraction of all others. In larger nuclei, the range of the nuclear force, shown for a single nucleon, is smaller than the size of the nucleus, but the Coulomb repulsion from all protons reaches all others. If the nucleus is large enough, the Coulomb repulsion can add to overcome the nuclear attraction.

There are some noticeable spikes on the BE/A graph, which represent particularly tightly bound nuclei. These spikes reveal further details of nuclear forces, such as confirming that closed-shell nuclei (those with magic numbers of protons or neutrons or both) are more tightly bound. The spikes also indicate that some nuclei with even numbers for Z and N , and with $Z = N$, are exceptionally tightly bound. This finding can be correlated with some of the cosmic abundances of the elements. The most common elements in the universe, as determined by observations of atomic spectra from outer space, are hydrogen, followed by ${}^4\text{He}$, with much smaller amounts of ${}^{12}\text{C}$ and other elements. It should be noted that the heavier elements are created in supernova explosions, while the lighter ones are produced by nuclear fusion during the normal life cycles of stars, as will be discussed in subsequent chapters. The most common elements have the

most tightly bound nuclei. It is also no accident that one of the most tightly bound light nuclei is ${}^4\text{He}$, emitted in α decay.

Example:

What Is BE/A for an Alpha Particle?

Calculate the binding energy per nucleon of ${}^4\text{He}$, the α particle.

Strategy

To find BE/A , we first find BE using the Equation

$\text{BE} = \{[Zm({}^1\text{H}) + Nm_n] - m({}^AX)\}c^2$ and then divide by A . This is straightforward once we have looked up the appropriate atomic masses in [Appendix A](#).

Solution

The binding energy for a nucleus is given by the equation

Equation:

$$\text{BE} = \{[Zm({}^1\text{H}) + Nm_n] - m({}^AX)\}c^2.$$

For ${}^4\text{He}$, we have $Z = N = 2$; thus,

Equation:

$$\text{BE} = \{[2m({}^1\text{H}) + 2m_n] - m({}^4\text{He})\}c^2.$$

[Appendix A](#) gives these masses as $m({}^4\text{He}) = 4.002602 \text{ u}$, $m({}^1\text{H}) = 1.007825 \text{ u}$, and $m_n = 1.008665 \text{ u}$. Thus,

Equation:

$$\text{BE} = (0.030378 \text{ u})c^2.$$

Noting that $1 \text{ u} = 931.5 \text{ MeV}/c^2$, we find

Equation:

$$\text{BE} = (0.030378)(931.5 \text{ MeV}/c^2)c^2 = 28.3 \text{ MeV}.$$

Since $A = 4$, we see that BE/A is this number divided by 4, or

Equation:

$$\text{BE}/A = 7.07 \text{ MeV/nucleon.}$$

Discussion

This is a large binding energy per nucleon compared with those for other low-mass nuclei, which have $\text{BE}/A \approx 3 \text{ MeV/nucleon}$. This indicates that ${}^4\text{He}$ is tightly bound compared with its neighbors on the chart of the nuclides. You can see the spike representing this value of BE/A for ${}^4\text{He}$ on the graph in [\[link\]](#). This is why ${}^4\text{He}$ is stable. Since ${}^4\text{He}$ is tightly bound, it has less mass than other $A = 4$ nuclei and, therefore, cannot spontaneously decay into them. The large binding energy also helps to explain why some nuclei undergo α decay. Smaller mass in the decay products can mean energy release, and such decays can be spontaneous. Further, it can happen that two protons and two neutrons in a nucleus can randomly find themselves together, experience the exceptionally large nuclear force that binds this combination, and act as a ${}^4\text{He}$ unit within the nucleus, at least for a while. In some cases, the ${}^4\text{He}$ escapes, and α decay has then taken place.

There is more to be learned from nuclear binding energies. The general trend in BE/A is fundamental to energy production in stars, and to fusion and fission energy sources on Earth, for example. This is one of the applications of nuclear physics covered in [Medical Applications of Nuclear Physics](#). The abundance of elements on Earth, in stars, and in the universe as a whole is related to the binding energy of nuclei and has implications for the continued expansion of the universe.

Problem-Solving Strategies

For Reaction And Binding Energies and Activity Calculations in Nuclear Physics

1. *Identify exactly what needs to be determined in the problem (identify the unknowns).* This will allow you to decide whether the energy of a decay or nuclear reaction is involved, for example, or whether the problem is primarily concerned with activity (rate of decay).

2. *Make a list of what is given or can be inferred from the problem as stated (identify the knowns).*
3. *For reaction and binding-energy problems, we use atomic rather than nuclear masses.* Since the masses of neutral atoms are used, you must count the number of electrons involved. If these do not balance (such as in β^+ decay), then an energy adjustment of 0.511 MeV per electron must be made. Also note that atomic masses may not be given in a problem; they can be found in tables.
4. *For problems involving activity, the relationship of activity to half-life, and the number of nuclei given in the equation $R = \frac{0.693N}{t_{1/2}}$ can be very useful.* Owing to the fact that number of nuclei is involved, you will also need to be familiar with moles and Avogadro's number.
5. *Perform the desired calculation; keep careful track of plus and minus signs as well as powers of 10.*
6. *Check the answer to see if it is reasonable: Does it make sense?*
Compare your results with worked examples and other information in the text. (Heeding the advice in Step 5 will also help you to be certain of your result.) You must understand the problem conceptually to be able to determine whether the numerical result is reasonable.

Note:**PhET Explorations: Nuclear Fission**

Start a chain reaction, or introduce non-radioactive isotopes to prevent one. Control energy production in a nuclear reactor!

<https://archive.cnx.org/specials/01caf0d0-116f-11e6-b891-abfdaa77b03b/nuclear-fission/#sim-one-nucleus>

Section Summary

- The binding energy (BE) of a nucleus is the energy needed to separate it into individual protons and neutrons. In terms of atomic masses,
Equation:

$$\text{BE} = \{[Zm(^1\text{H}) + Nm_n] - m(^A\text{X})\}c^2,$$

where $m(^1\text{H})$ is the mass of a hydrogen atom, $m(^A\text{X})$ is the atomic mass of the nuclide, and m_n is the mass of a neutron. Patterns in the binding energy per nucleon, BE/A , reveal details of the nuclear force. The larger the BE/A , the more stable the nucleus.

Conceptual Questions

Exercise:

Problem:

Why is the number of neutrons greater than the number of protons in stable nuclei having A greater than about 40, and why is this effect more pronounced for the heaviest nuclei?

Problems & Exercises

Exercise:

Problem:

^2H is a loosely bound isotope of hydrogen. Called deuterium or heavy hydrogen, it is stable but relatively rare—it is 0.015% of natural hydrogen. Note that deuterium has $Z = N$, which should tend to make it more tightly bound, but both are odd numbers. Calculate BE/A , the binding energy per nucleon, for ^2H and compare it with the approximate value obtained from the graph in [\[link\]](#).

Solution:

1.112 MeV, consistent with graph

Exercise:

Problem:

^{56}Fe is among the most tightly bound of all nuclides. It is more than 90% of natural iron. Note that ^{56}Fe has even numbers of both protons and neutrons. Calculate BE/A , the binding energy per nucleon, for ^{56}Fe and compare it with the approximate value obtained from the graph in [\[link\]](#).

Exercise:**Problem:**

^{209}Bi is the heaviest stable nuclide, and its BE/A is low compared with medium-mass nuclides. Calculate BE/A , the binding energy per nucleon, for ^{209}Bi and compare it with the approximate value obtained from the graph in [\[link\]](#).

Solution:

7.848 MeV, consistent with graph

Exercise:**Problem:**

(a) Calculate BE/A for ^{235}U , the rarer of the two most common uranium isotopes. (b) Calculate BE/A for ^{238}U . (Most of uranium is ^{238}U .) Note that ^{238}U has even numbers of both protons and neutrons. Is the BE/A of ^{238}U significantly different from that of ^{235}U ?

Exercise:**Problem:**

(a) Calculate BE/A for ^{12}C . Stable and relatively tightly bound, this nuclide is most of natural carbon. (b) Calculate BE/A for ^{14}C . Is the difference in BE/A between ^{12}C and ^{14}C significant? One is stable and common, and the other is unstable and rare.

Solution:

(a) 7.680 MeV, consistent with graph

(b) 7.520 MeV, consistent with graph. Not significantly different from value for ^{12}C , but sufficiently lower to allow decay into another nuclide that is more tightly bound.

Exercise:

Problem:

The fact that BE/A is greatest for A near 60 implies that the range of the nuclear force is about the diameter of such nuclides. (a) Calculate the diameter of an $A = 60$ nucleus. (b) Compare BE/A for ^{58}Ni and ^{90}Sr . The first is one of the most tightly bound nuclides, while the second is larger and less tightly bound.

Exercise:

Problem:

The purpose of this problem is to show in three ways that the binding energy of the electron in a hydrogen atom is negligible compared with the masses of the proton and electron. (a) Calculate the mass equivalent in u of the 13.6-eV binding energy of an electron in a hydrogen atom, and compare this with the mass of the hydrogen atom obtained from [Appendix A](#). (b) Subtract the mass of the proton given in [\[link\]](#) from the mass of the hydrogen atom given in [Appendix A](#). You will find the difference is equal to the electron's mass to three digits, implying the binding energy is small in comparison. (c) Take the ratio of the binding energy of the electron (13.6 eV) to the energy equivalent of the electron's mass (0.511 MeV). (d) Discuss how your answers confirm the stated purpose of this problem.

Solution:

(a) 1.46×10^{-8} u vs. 1.007825 u for ^1H

(b) 0.000549 u

(c) 2.66×10^{-5}

Exercise:

Problem: Unreasonable Results

A particle physicist discovers a neutral particle with a mass of 2.02733 u that he assumes is two neutrons bound together. (a) Find the binding energy. (b) What is unreasonable about this result? (c) What assumptions are unreasonable or inconsistent?

Solution:

(a) -9.315 MeV

(b) The negative binding energy implies an unbound system.

(c) This assumption that it is two bound neutrons is incorrect.

Glossary

binding energy

the energy needed to separate nucleus into individual protons and neutrons

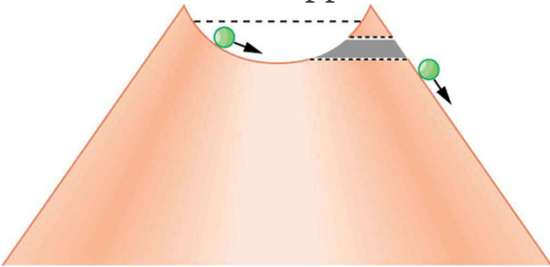
binding energy per nucleon

the binding energy calculated per nucleon; it reveals the details of the nuclear force—larger the BE/A , the more stable the nucleus

Tunneling

- Define and discuss tunneling.
- Define potential barrier.
- Explain quantum tunneling.

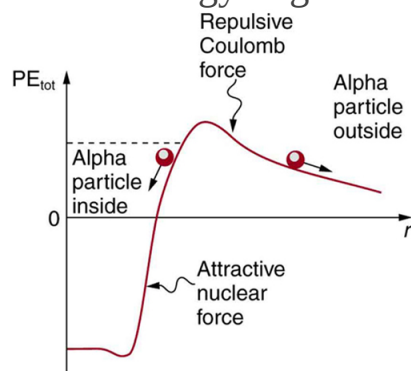
Protons and neutrons are *bound* inside nuclei, that means energy must be supplied to break them away. The situation is analogous to a marble in a bowl that can roll around but lacks the energy to get over the rim. It is bound inside the bowl (see [\[link\]](#)). If the marble could get over the rim, it would gain kinetic energy by rolling down outside. However classically, if the marble does not have enough kinetic energy to get over the rim, it remains forever trapped in its well.



The marble in this semicircular bowl at the top of a volcano has enough kinetic energy to get to the altitude of the dashed line, but not enough to get over the rim, so that it is trapped forever. If it could find a tunnel through the barrier, it would escape, roll downhill, and gain kinetic energy.

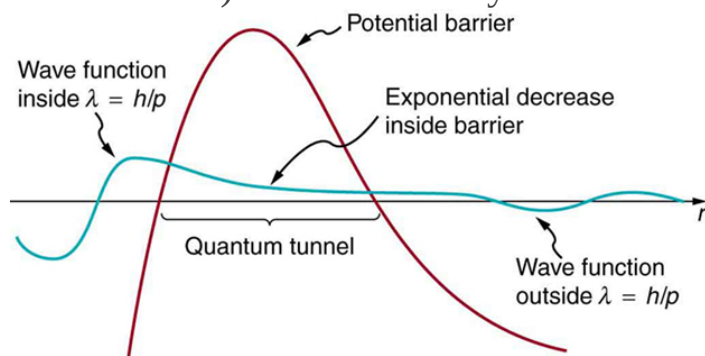
In a nucleus, the attractive nuclear potential is analogous to the bowl at the top of a volcano (where the “volcano” refers only to the shape). Protons and neutrons have kinetic energy, but it is about 8 MeV less than that needed to get out (see [\[link\]](#)). That is, they are bound by an average of 8 MeV per

nucleon. The slope of the hill outside the bowl is analogous to the repulsive Coulomb potential for a nucleus, such as for an α particle outside a positive nucleus. In α decay, two protons and two neutrons spontaneously break away as a ${}^4\text{He}$ unit. Yet the protons and neutrons do not have enough kinetic energy to get over the rim. So how does the α particle get out?



Nucleons within an atomic nucleus are bound or trapped by the attractive nuclear force, as shown in this simplified potential energy curve. An α particle outside the range of the nuclear force feels the repulsive Coulomb force. The α particle inside the nucleus does not have enough kinetic energy to get over the rim, yet it does manage to get out by quantum mechanical tunneling.

The answer was supplied in 1928 by the Russian physicist George Gamow (1904–1968). The α particle tunnels through a region of space it is forbidden to be in, and it comes out of the side of the nucleus. Like an electron making a transition between orbits around an atom, it travels from one point to another without ever having been in between. [\[link\]](#) indicates how this works. The wave function of a quantum mechanical particle varies smoothly, going from within an atomic nucleus (on one side of a potential energy barrier) to outside the nucleus (on the other side of the potential energy barrier). Inside the barrier, the wave function does not become zero but decreases exponentially, and we do not observe the particle inside the barrier. The probability of finding a particle is related to the square of its wave function, and so there is a small probability of finding the particle outside the barrier, which implies that the particle can tunnel through the barrier. This process is called **barrier penetration** or **quantum mechanical tunneling**. This concept was developed in theory by J. Robert Oppenheimer (who led the development of the first nuclear bombs during World War II) and was used by Gamow and others to describe α decay.

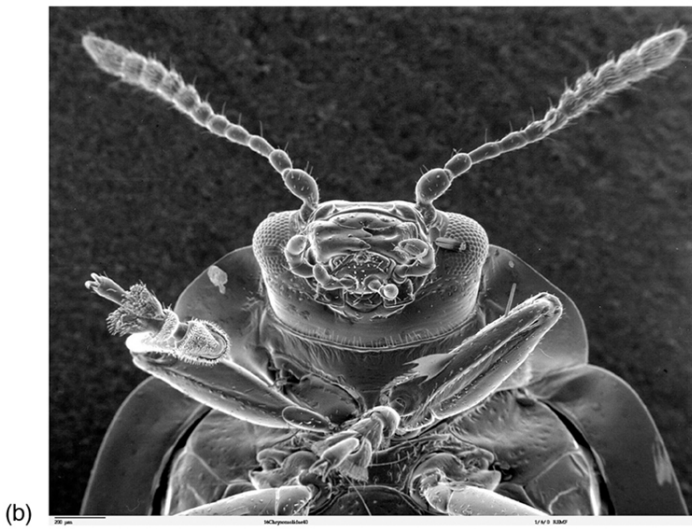
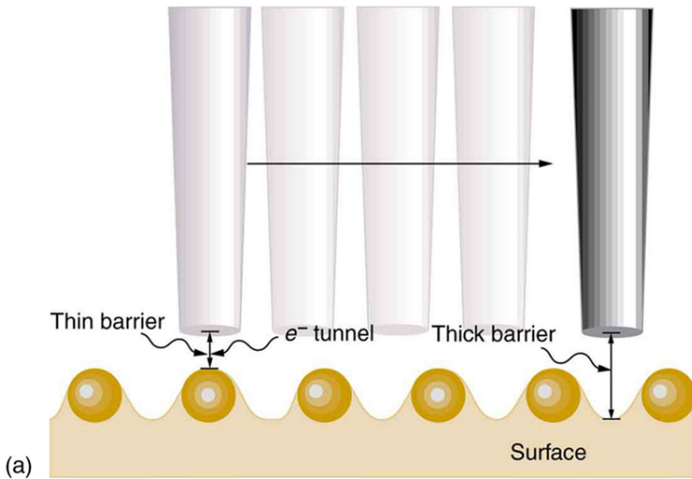


The wave function representing a quantum mechanical particle must vary smoothly, going from within the nucleus (to the left of the barrier) to outside the nucleus (to the right of the barrier). Inside the barrier, the wave function does not abruptly become zero; rather, it decreases exponentially. Outside the barrier, the

wave function is small but finite, and there it smoothly becomes sinusoidal. Owing to the fact that there is a small probability of finding the particle outside the barrier, the particle can tunnel through the barrier.

Good ideas explain more than one thing. In addition to qualitatively explaining how the four nucleons in an α particle can get out of the nucleus, the detailed theory also explains quantitatively the half-life of various nuclei that undergo α decay. This description is what Gamow and others devised, and it works for α decay half-lives that vary by 17 orders of magnitude. Experiments have shown that the more energetic the α decay of a particular nuclide is, the shorter is its half-life. **Tunneling** explains this in the following manner: For the decay to be more energetic, the nucleons must have more energy in the nucleus and should be able to ascend a little closer to the rim. The barrier is therefore not as thick for more energetic decay, and the exponential decrease of the wave function inside the barrier is not as great. Thus the probability of finding the particle outside the barrier is greater, and the half-life is shorter.

Tunneling as an effect also occurs in quantum mechanical systems other than nuclei. Electrons trapped in solids can tunnel from one object to another if the barrier between the objects is thin enough. The process is the same in principle as described for α decay. It is far more likely for a thin barrier than a thick one. Scanning tunneling electron microscopes function on this principle. The current of electrons that travels between a probe and a sample tunnels through a barrier and is very sensitive to its thickness, allowing detection of individual atoms as shown in [\[link\]](#).



(a) A scanning tunneling electron microscope can detect extremely small variations in dimensions, such as individual atoms. Electrons tunnel quantum mechanically between the probe and the sample. The probability of tunneling is extremely sensitive to barrier thickness, so that the electron current is a sensitive indicator of surface features. (b) Head and mouthparts of *Coleoptera Chrysomelidea* as seen through an electron microscope (credit: Louisa Howard, Dartmouth College)

Note:

PhET Explorations: Quantum Tunneling and Wave Packets

Watch quantum "particles" tunnel through barriers. Explore the properties of the wave functions that describe these particles. [Click to open media in new browser.](#)

Section Summary

- Tunneling is a quantum mechanical process of potential energy barrier penetration. The concept was first applied to explain α decay, but tunneling is found to occur in other quantum mechanical systems.

Conceptual Questions

Exercise:**Problem:**

A physics student caught breaking conservation laws is imprisoned. She leans against the cell wall hoping to tunnel out quantum mechanically. Explain why her chances are negligible. (This is so in any classical situation.)

Exercise:**Problem:**

When a nucleus α decays, does the α particle move continuously from inside the nucleus to outside? That is, does it travel each point along an imaginary line from inside to out? Explain.

Problems-Exercises

Exercise:

Problem:

Derive an approximate relationship between the energy of α decay and half-life using the following data. It may be useful to graph the log of $t_{1/2}$ against E_α to find some straight-line relationship.

Nuclide	E_α (MeV)	$t_{1/2}$
^{216}Ra	9.5	0.18 μs
^{194}Po	7.0	0.7 s
^{240}Cm	6.4	27 d
^{226}Ra	4.91	1600 y
^{232}Th	4.1	1.4×10^{10} y

Energy and Half-Life for α Decay

Exercise:

Problem: Integrated Concepts

A 2.00-T magnetic field is applied perpendicular to the path of charged particles in a bubble chamber. What is the radius of curvature of the path of a 10 MeV proton in this field? Neglect any slowing along its path.

Solution:

22.8 cm

Exercise:**Problem:**

(a) Write the decay equation for the α decay of ^{235}U . (b) What energy is released in this decay? The mass of the daughter nuclide is 231.036298 u. (c) Assuming the residual nucleus is formed in its ground state, how much energy goes to the α particle?

Solution:

(b) 4.679 MeV

(c) 4.599 MeV

Exercise:**Problem: Unreasonable Results**

The relatively scarce naturally occurring calcium isotope ^{48}Ca has a half-life of about 2×10^{16} y. (a) A small sample of this isotope is labeled as having an activity of 1.0 Ci. What is the mass of the ^{48}Ca in the sample? (b) What is unreasonable about this result? (c) What assumption is responsible?

Exercise:**Problem: Unreasonable Results**

A physicist scatters γ rays from a substance and sees evidence of a nucleus 7.5×10^{-13} m in radius. (a) Find the atomic mass of such a nucleus. (b) What is unreasonable about this result? (c) What is unreasonable about the assumption?

Solution:

a) $2.4 \times 10^8 \text{ u}$

(b) The greatest known atomic masses are about 260. This result found in (a) is extremely large.

(c) The assumed radius is much too large to be reasonable.

Exercise:**Problem: Unreasonable Results**

A frazzled theoretical physicist reckons that all conservation laws are obeyed in the decay of a proton into a neutron, positron, and neutrino (as in β^+ decay of a nucleus) and sends a paper to a journal to announce the reaction as a possible end of the universe due to the spontaneous decay of protons. (a) What energy is released in this decay? (b) What is unreasonable about this result? (c) What assumption is responsible?

Solution:

(a) -1.805 MeV

(b) Negative energy implies energy input is necessary and the reaction cannot be spontaneous.

(c) Although all conservation laws are obeyed, energy must be supplied, so the assumption of spontaneous decay is incorrect.

Exercise:**Problem: Construct Your Own Problem**

Consider the decay of radioactive substances in the Earth's interior. The energy emitted is converted to thermal energy that reaches the earth's surface and is radiated away into cold dark space. Construct a

problem in which you estimate the activity in a cubic meter of earth rock? And then calculate the power generated. Calculate how much power must cross each square meter of the Earth's surface if the power is dissipated at the same rate as it is generated. Among the things to consider are the activity per cubic meter, the energy per decay, and the size of the Earth.

Glossary

barrier penetration

quantum mechanical effect whereby a particle has a nonzero probability to cross through a potential energy barrier despite not having sufficient energy to pass over the barrier; also called quantum mechanical tunneling

quantum mechanical tunneling

quantum mechanical effect whereby a particle has a nonzero probability to cross through a potential energy barrier despite not having sufficient energy to pass over the barrier; also called barrier penetration

tunneling

a quantum mechanical process of potential energy barrier penetration